Ion Purcaru

Informație și corelație

Editura Științifică și Enciclopedică

Scanned with OKEN Scanner

Ion Purcaru

INFORMAȚIE ȘI CORELAȚIE

EDITURA ȘTIINȚIFICĂ ȘI ENCICLOPEDICĂ
București 1988

Scanned with OKEN Scanner

PREFAȚĂ

Apărută pe un așa zis „teren al nimănui", ai cărui „vecini" erau matematica, fizica, electronica, electrotehnica etc., teoria matematică a informației este relativ tînără în raport cu alte ramuri ale științei, anul 1988 marcîndu-i cea de a patruzecea aniversare. Deși tînără, ea s-a dezvoltat destul de repede devenind un domeniu de cercetare științifică de sine stătător, în conexiune cu alte domenii ale științei dintre cele mai diverse.

Elaborată și cu gîndul la acest jubileu al teoriei informației, lucrarea „Informație și corelație" are ca scop utilizarea conceptului de entropie pentru definirea și studierea unei clase de măsuri ale dependenței probabiliste, denumite prin urmare măsuri entropice ale dependenței.

Prima parte a lucrării constituie o sinteză a principalelor rezultate asupra măsurării cantității și utilității informației, bazate pe conceptele de entropie Shannon (1948) și entropie ponderată Guiașu (1968). A doua parte este de asemenea o sinteză, dar a principalelor rezultate asupra măsurării dependenței probabiliste, subordonate în prezentare, pentru unitatea acestei părți, conceptului de bună măsură a dependenței introdus de A. Rényi (1959). În această parte a lucrării se acordă o importanță deosebită măsurilor entropice ale dependenței pentru a căror prezentare sînt utilizate îndeosebi rezultatele din paragrafele 3.2, 3.3 și 4.2 ale primei părți, ceea ce asigură și o conexiune între primele două părți ale lucrării. Prin intermediul măsurilor entropice ale dependenței se asigură și legătura cu cea de a treia parte, în care, pornind de la conceptul de organizare W—G (1969) sînt prezentate posibilități de aplicare a acestor măsuri în studiul și analiza unor probleme din teoria matematică a sistemelor aleatoare, cu unele referiri la sisteme de transmitere a informației sau la sisteme de strategii aleatoare.

Scanned with OKEN Scanner

În ultimul capitol se reia noțiunea de informație, cu care începe lucrarea, prin prisma conexiunii dintre informație și probabilitate, făcîndu-se anumite considerații asupra celor două concepte ce stau, rînd pe rînd, la baza tuturor rezultatelor prezentate în lucrare.

Ca urmare a modului în care este concepută, lucrarea se adresează unei categorii largi de cititori, matematicieni, ingineri, ciberneticieni, economiști sau de alte specialități, cu preocupări asupra teoriei matematice a informației sau interesați de unele probleme sau aplicații ale acesteia.

Aș dori ca pe această cale să aduc mulțumirile mele respectuoase tuturor celor de la care învățînd, de-a lungul anilor, am ajuns să elaborez prezenta lucrare. De asemenea, și nu în ultimul rînd, doresc să mulțumesc sincer Editurii Științifice și Enciclopedice care mi-a facilitat publicarea acestei lucrări.

Desigur, ca orice rezultat al cercetării, lucrarea „Informație și corelație" constituie o sinteză de preocupări, idei și rezultate din domeniul pe care l-am abordat, existînd, ca întotdeauna loc pentru mai mult și mai bine. Din acest motiv doresc să mulțumesc anticipat cititorilor lucrării, care vor găsi de cuviință, să facă observații asupra formei și conținutului acesteia, de care voi ține cont la o eventuală reeditare și mai ales în activitatea mea de cercetare sau în cea didactică.

AUTORUL

I. MĂSURA INFORMAȚIEI

1. Apariția teoriei informației

Conceptul de informație a stîrnit numeroase discuții, adeseori contradictorii, căci informația, în sensul ei cel mai general, există de cînd lumea. Ajungîndu-se însă la o teorie a informației, s-a pus în mod firesc întrebarea „ce este informația?" S-au dat numeroase încercări de răspunsuri, pe care însă dicționarele nu le-au putut sintetiza într-o definiție unitară. Putem vorbi de informații obținute pe cale auditivă, vizuală sau de altă natură, în fiecare caz existînd practic o infinitate de situații ce pot fi exemplificate. De fiecare dată însă, dacă încercăm să definim informația, sîntem nevoiți să apelăm la exemple de informații. Acest fapt ne conduce cu gîndul la conceptul de mulțime și la ideea că și informația ar putea, mai mult chiar ar trebui, să fie considerată ca un concept primar. La fel de firesc cu întrebarea „ce este informația"?, ca noțiune generală, s-a pus și altă chestiune „ce reprezintă informația?" ca alcătuire. Cu alte cuvinte ce este ea, materie sau energie? Controversele au existat și există și asupra acestei chestiuni, oameni de știință din cele mai diverse domenii încercînd elaborarea unui răspuns unanim acceptabil. Savantul american Norbert Wiener (1894—1964) spunea: „Creierul mecanic nu secretă gîndire, precum ficatul fiere, cum afirmau materialiștii din trecut; de asemenea el nu o degajă sub formă de energie cum își degajă mușchii energia. Informația este informație, nici materie și nici energie. Nici un materialism care nu admite aceasta nu poate exista în prezent". Este foarte adevărat că transmiterea unei informații presupune existența unui minim de energie, dar cantitatea de informație nu depinde de valoarea acestui minim, ci numai calitatea informației eventual.

Scanned with OKEN Scanner

Un profesor poate spune elevilor săi același lucru cu o voce tunătoare, sau cu o voce calmă, natural dacă este auzit. De asemenea, profesorul nu pierde informațiile pe care le predă elevilor săi. Mai mult chiar, aceleași informații pot fi la fel de bine transmise de un coleg al acestuia.

Avem deci un exemplu clar de situații în care informația nu poate fi identificată nici cu materia și nici cu energia. Astfel de exemple există pretutindeni. Ele au condus însă și la concluzii filozofice dintre cele mai diverse privind natura informației, aceasta fiind chiar desprinsă de lumea materială.

Noțiunea de informație este intim legată de transmiterea ei. În ipoteza unității materiale a lumii, transmiterea informației nu poate depăși niciodată limitele realității. Ea leagă între ele, într-un anumit fel, diverse elemente definitorii ale ei : mașină cu mașină, om cu om, mașină cu om, organism cu organism, conștiința unui om de conștiința altuia etc., ceea ce conduce la ideea că informația nu poate apare și nici exista în afara sferei materiale. Informația apare deci ca un raport special între procese materiale, raport care nu există în afara acestora. Ea poate fi considerată ca o însușire a materiei, dar nu poate fi identificată cu ea.

În ciuda tuturor controverselor privind conceptul de informație, teoria informației a apărut și s-a dezvoltat foarte repede, fiind însă destul de tînără, căci în curînd va împlini 40 de ani. Dat fiind aparatul matematic care stă la baza ei, teoria informației ar putea fi considerată ca ramură a teoriei probabilităților și statisticii matematice. Dacă ne gîndim la problematica generală a teoriei comunicațiilor care a condus la apariția ei, teoria informației poate fi considerată și ramură a altor teorii. O privire atentă și riguroasă asupra teoriei informației poate însă conduce la concluzia considerării acesteia ca un domeniu științific de sine stătător, chiar dacă a apărut pe „un teren al nimănui", ai cărui „vecini" erau matematica, fizica, electronica, electrotehnica etc.... Ea nu a apărut întîmplător tocmai pe acest „teren al nimănui", de altfel foarte fertil, unde lumea era preocupată de problemele generale ale comenzii și comunicației ce aveau să conducă la fundamentarea ciberneticii. Și nu a apărut nici brusc. Este o apariție relativ lentă dar spectaculoasă, reprezentînd o sinteză de preocupări și rezultate diverse, aparent nelegate între ele, urmată de o generare de idei creatoare de alte domenii de sine stătătoare ale cercetării fundamentale sau aplicative.

Scanned with OKEN Scanner

În anul 1838 a apărut pentru prima dată termenul de cibernetică, ca propunere a savantului francez A.M. Ampère (1775—1836), într-o încercare de clasificare a științelor, pentru acea parte a economiei politice care se ocupa de arta guvernării. Preocupat de teoria căldurii și dezvoltînd lucrările lui S. Carnot (1796 — 1832), și B. P. Clapeyron (1799—1864), fizicianul german R.E. Clausius (1822—1888), a publicat în 1850 articolul „Despre forța motoare a căldurii" în care punea în evidență procesul de degradare a energiei și prin urmare principiul al doilea al termodinamicii. În 1856 fizicianul scoțian J.C. Maxwell (1831—1879) a prezentat o comunicare în ședința Academiei britanice din Aberdeen, în care pentru prima dată se împleteau principiile mecanicii cu noțiunile teoriei probabilității și se stabilea legea de repartiție a vitezelor particulelor ca urmare a ciocnirilor dintre ele, demonstrîndu-se că echilibrul unui sistem fizic se realizează numai în urma unui mare număr de interacțiuni dezordonate ale particulelor componente ale sistemului. În 1859 Clausius a dat o formulare matematică principiului al doilea al termodinamicii introducînd pentru aceasta o funcție pe care a numit-o entropie, a cărei variație dS este raportul dintre cantitatea infinitezimală de căldură dQ pe care o capătă un sistem aflat într-un proces reversibil (proces care la inversarea totală a condițiilor evoluează ca și procesul direct dar în sens invers) și temperatura absolută T la care s-a făcut transmiterea căldurii (dS=dQ/T). Cu ajutorul entropiei, principiul al doilea al termodinamicii afirmă că într-un sistem, entropia sau rămîne staționară, sau evoluează în sens crescător. În 1868 Maxwell a publicat prima analiză teoretică a adaptabilității în care a expus pentru prima dată principiul unui dispozitiv cu feedback (celebrul exemplu cu demonul lui Maxwell), pentru ca în 1871 să fie tot primul care a apropiat noțiunea de informație de noțiunile de bază ale energeticii. În același an 1871, fizicianul austriac L. Boltzmann (1844—1906) a demonstrat legea repartiției vitezelor dată de Maxwell în 1856 și pornind de la ea, a dedus principiul echipartiției energiei, iar în 1877 a reușit pentru prima dată să interpreteze principiul al doilea al termodinamicii ca o lege probabilistică stabilind pentru entropia unui sistem fizic formula $S = -k \sum_{i=1}^{n} p_i \log p_i$ unde k este constanta lui Boltzmann, iar p_i este probabilitatea ca sistemul să se afle în starea i, $1 \leqslant i \leqslant n$. Sub forma dată de Boltzmann, entropia exprimă limpede gradul de dezordine al sistemului fizic.

Scanned with OKEN Scanner

La o jumătate de secol după acest rezultat, inginerul electrician american R.W.L. Hartley stabileşte că transmiterea unui mesaj într-un timp dat este măsurabilă şi limitată şi propune în 1927 ca măsură a gradului de nedeterminare al unui experiment cu n rezultate posibile mărimea $\log n$. În 1935 statisticianul englez R.S. Fisher (1890—1962), fondator al celei mai mari părţi din statistica matematică, arată că ideile teoriei informaţiei izvorăsc din noţiunea de dezordine sau entropie din termodinamică şi mecanica statistică. El a legat teoria informaţiei de teoria inferenţei statistice, ocazie cu care a introdus şi conceptul de „statistici suficiente" ca reprezentînd acele eşantioane care să conţină maximum de informaţie despre parametrii studiaţi. Doi ani mai tîrziu, în 1937, inginerul electrotehnist şi matematicianul american C.E. Shannon demonstrează că circuitele cu comutare automată se supun teoriei algebrei Boole. În 1938 fiziologul american A. Rosenblueth conducea o serie de întruniri lunare asupra problemelor de metodologie ştiinţifică, la care luau parte specialişti din domenii diverse, subliniind necesitatea colectivelor ştiinţifice interdisciplinare. La aceste dezbateri a apărut şi apoi a devenit deosebit de activ şi matematicianul N. Wiener. Colectivul condus de cei doi ajunge să fondeze cibernetica în 1942, expunînd deja unele idei la o sesiune ştiinţifică la New York, fără a utiliza acest termen. Colectivul se măreşte în 1943—1944 cu alte personalităţi, între care matematicianul J. von Neumann (1903—1957). Din iniţiativa lui Wiener şi von Neumann, la sfîrşitul lui 1943 şi începutul lui 1944 s-a organizat la Princeton o întrunire a tuturor celor preocupaţi de problemele de comunicaţie şi conducere, perioadă în care Wiener reia pentru prima oară conceptul de feedback introdus de Maxwell în 1868. Şi iată cum acel „teren al nimănui" îşi găsea „stăpîni"! Munca lor duce la apariţia în 1948 a două lucrări fundamentale în domeniul comunicaţiei şi conducerii şi anume: „Cibernetica sau ştiinţa comenzii şi comunicării la fiinţe şi maşini" a lui N. Wiener şi „O teorie matematică a comunicaţiei" a lui C. E. Shannon. Lucrarea lui C. E. Shannon fundamenta din punct de vedere matematic teoria informaţiei, creînd tocmai acel aparat matematic care lipsea din lucrarea lui N. Wiener. Îmbinarea în mod armonios a celor două lucrări a constituit fundamentul noii ştiinţe, cibernetica, „recoltă" de pe acel roditor „teren al nimănui" constituit de domeniul comunicaţiei şi conducerii, pe care lucraseră cu multă abnegaţie specialişti în matematică, fizică, medicină,

electrotehnică și alte domenii ! Idei diverse, elaborate timp de un secol, se contopeau acum cu măiestrie, generînd un domeniu nou de cercetare și ilustrînd în mod concret o remarcă a lui Mircea Malița potrivit căreia : „Drumul științei este pacient, lent și cumulativ. Știința în toate vremurile este fructul acelei grădini pe care fiii au săpat-o de nenumărate ori făcînd-o roditoare, după ce tatăl lor le spusese povestea că în adîncul ei stă o comoară".

Deși abia născută, teoria informației cunoaște o dezvoltare remarcabilă. În 1950 se organizează la Londra primul simpozion de teoria matematică a informației, iar în 1953 apare revista „IRE The Transactions of Information Theory". În același an 1953, deci la cinci ani de la apariția lucrării lui Shannon, erau enumerate 979 de lucrări de teoria informației, dintre care doar 104 erau elaborate pînă în 1948 ! Dezvoltarea este remarcabilă cantitativ și calitativ. În 1953 matematicianul sovietic A. I. Hincin (1894—1959) reia teoria lui Shannon dînd demonstrații riguroase principalelor rezultate. În 1956 Fisher, Shannon și Wiener subliniază caracterul statistico-matematic al teoriei informației. În 1958 apare cea de a doua revistă a noii discipline, „Information and Control", care publică în mare majoritate articole de teoria informației. Autorul axiomaticii teoriei probabilității, matematicianul sovietic A. N. Kolmogorov, intervine și el asupra rigurozității aparatului matematic al teoriei informației subliniind în 1958 că „cibernetica este știința modurilor de recepție, păstrare, transformare și folosire a informației unor mașini, organisme vii și reuniuni ale lor", iar în 1959 că expresia matematică a cantității de informație „este formal identică cu expresia entropiei din fizică... Ar fi însă o exagerare să considerăm că teoriile fizice legate de noțiunea de entropie ar conține în sine de-a gata elementele teoriei informației". Nu este vorba în apariția ciberneticii și a teoriei informației doar de o simplă analogie, între procesul de transmitere a căldurii și acela de transmitere a informației, dar analogia a jucat aici un rol primordial, ilustrîndu-se odată în plus, dar nu prima și nici ultima dată, spusele matematicianului polonez Ștefan Banach (1892—1945) : „Matematician mare este cel care știe să găsească analogia între o teoremă și alta ; mai bun matematician este cel care vede analogiile între demonstrații ; și mai ridicat este cel care observă analogii între teorii ; ne putem închipui chiar pe cineva care să vadă analogiile între analogii".

Scanned with OKEN Scanner

Se făcuse deja o analogie între două teorii, ca modele generale a două domenii total diferite, care evoluaseră în paralel multă vreme, aparent fără nici o legătură. Conceptul de informaţie stabilea formal o punte între ele, fără a le identifica.

Într-un deceniu de la apariţia lucrării lui Shannon aparatul matematic al teoriei informaţiei era riguros pus la punct. De acum începe dezvoltarea teoriei matematice a informaţiei cu precădere prin aplicaţiile sale în cele mai diferite domenii precum teoria codurilor, teoria jocurilor,teoria automatelor, teoria sistemelor şi altele, ele însăşi cîndva „terenuri ale nimănui", iar astăzi domenii de sine stătătoare.

Cîte lucrări s-au elaborat în domeniul teoriei informaţiei pînă în prezent, este greu de precizat, dar oricum sînt cîteva mii. S-ar putea pune întebarea dacă mai vorbim de teoria informaţiei ca disciplină, sau de altceva, mult mai solid şi mai puternic, clădit pe scheletul teoriei matematice a informaţiei. N-am greşi zicîndu-i „Cercetări informaţionale"!

2. Problema măsurării informaţiei

2.1. Consideraţii generale

Deşi noţiunea de informaţie are o mare generalitate şi cu toate că nu i s-a putut da încă o definiţie unitară, fiind considerată un concept primar ca şi noţiunea de mulţime, s-a pus, în mod natural, problema măsurării ei.

Dacă informaţia, în cazul unui sistem oarecare, tehnic, social, biologic sau de altă natură, poate fi considerată ca un mesaj despre evenimente care au avut, au sau vor avea loc în interiorul cît şi în exteriorul acestui sistem, atunci, luînd în consideraţie două mesaje despre un acelaşi eveniment, apar în mod firesc trei întrebări:

a) care dintre cele două mesaje este mai mare?

b) care dintre cele două mesaje este mai important şi deci mai util?

c) cît de complet este un anumit mesaj?

Şi într-un caz şi în celălalt comparaţia dintre cele două mesaje impune existenţa unei măsuri a informaţiei pe care o vom obţine în urma realizării unui eveniment, informaţie ce înlătură nedeterminarea pe care o aveam înainte de efec-

Scanned with OKEN Scanner

tuarea experimentului ce implică realizarea sau nerealizarea evenimentului amintit. Între informație și nedeterminare există o legătură strînsă, informația dovedindu-se a fi informație în adevăratul sens al cuvîntului atunci și numai atunci cînd ea înlătură o anumită nedeterminare. Cu cît nedeterminarea de la începutul experimentului este mai mare, cu atît este mai mare informația ce se obține după efectuarea experimentului. Se poate spune că informația înlocuiește o nedeterminare și dacă aceasta din urmă poate fi măsurată, atunci și informația poate fi măsurată. Cantitativ, cele două concepte pot fi identificate și deci măsurate în același mod, dar trebuie privite separat în ceea ce privește sensul de variație și semnificația fiecăruia.

Cele trei întrebări amintite privesc atît cantitatea de informație furnizată de realizarea unui anumit eveniment, cît și utilitatea informației obținute. Ele completează în fapt întrebarea privind existența informației. Am putea spune pe scurt: „ce este și cum se măsoară informația"? S-au remarcat diverse răspunsuri. În ceea ce privește completitudinea informației, în domeniul său de preocupări, statistica matematică, Fisher a dat deja un răspuns în 1935 prin introducerea statisticilor suficiente, reluînd și dezvoltîndu-și unele idei și rezultate din 1925. O primă soluție privind măsurarea cantității de informație, cu meritul istoric cuvenit, a avut-o Hartley în 1927, propunînd drept măsură a cantității de informație conținută de un experiment cu n rezultate expresia $\log n$. Soluția definitivă pentru această chestiune avea să o dea Shannon în 1948. Doctor în matematici și magistru în electrotehnică la vîrsta de 24 de ani, C. E. Shannon, după ce demonstrează că schemele cu contacte și relee pot fi ușor studiate cu ajutorul aparatului algebrei Boole, are ideea analogiei dintre transmiterea căldurii într-un sistem fizic, descrisă riguros de Boltzmann în 1877, și transmiterea informației într-un sistem oarecare destinat acestui proces. Această idee îl conduce în 1948 la soluționarea problemei măsurării cantității de informație furnizată de realizarea unui experiment cu n rezultate posibile, devenind astfel la 32 de ani fondatorul de necontestat al teoriei matematice a informației, pe care o va dezvolta ulterior cu precădere în domeniul transmiterii informației în medii cu sau fără perturbații. Idei și soluții pentru această problemă au existat și după 1948, dar rezultatul lui Shannon a rămas ca un nucleu în jurul căruia s-a dezvoltat teoria cantitativă a informației. Cu privire la măsurarea utilității informației

Scanned with OKEN Scanner

se cuvine a menționa preocupările lui S. Guiașu din 1968 prin introducerea entropiei ponderate, analogă conceptului de utilitate introdus de J. von Neumann și O. Morgenstern în 1944, el însuși supus atenției matematicienilor și tuturor celor preocupați de teoria jocurilor încă din 1730 de către Daniel Bernoulli (1700—1782). Așadar s-au încercat soluții pentru toate cele trei întrebări ce apar ca naturale. Răspunsul cel mai bine pus la punct, pînă în prezent, rămîne cel legat de măsura cantității de informație dat de Shannon.

2.2. Axiomele măsurii cantității de informație

În ceea ce urmează vom presupune cunoscute conceptele de eveniment, cîmp de evenimente, probabilitate, cîmp de probabilitate, variabilă aleatoare, vector aleator și celelalte noțiuni de teoria probabilităților expuse riguros în lucrările lui O. Onicescu, Gh. Mihoc, G. Ciucu, I. Cuculescu, G. Sâmboan, M. Iosifescu, V. Craiu și alții.

Să considerăm așadar un cîmp de probabilitate $\{\Omega, \mathcal{K}, P\}$. Realizarea unui eveniment arbitrar $A \in \mathcal{K}$ implică obținerea unei cantități de informație pe care o vom nota cu $I(A)$.

Definiția 2.2.1. Se numește măsură a cantității de informație obținută ca urmare a realizării unui eveniment $A \in \mathcal{K}$ orice funcție $I : \mathcal{K} \to [0, \infty)$ care satisface în mod natural axiomele :

A.1 (Axioma existenței). $I(A)$ există pentru orice $A \in \mathcal{K}$

A.2 (Axioma valorilor extreme). Informația furnizată de un eveniment sigur este zero, iar cea furnizată de un eveniment imposibil infinită, adică

$$I(\Omega) = 0 \text{ și } I(\emptyset) = \infty \tag{2.2.1}$$

A.3 (Axioma de monotonie). Dacă evenimentul A implică realizarea evenimentului B atunci informația furnizată de realizarea lui A este cel puțin egală cu cea furnizată de realizarea lui B, adică

$$A \subseteq B \Rightarrow I(A) \geqslant I(B), \quad A, B \in \mathcal{K} \tag{2.2.2}$$

A.4 (Axioma reuniunii). Dacă $f : [0, \infty) \to \mathbb{R}$ este convexă și strict descrescătoare,

$$I(A \cup B) = f[f^{-1}(I(A)) + f^{-1}(I(B))]$$
$$(\forall)\ A, B \in \mathcal{K}, \quad A \cap B = \emptyset \tag{2.2.3}$$

Scanned with OKEN Scanner

A.5 (Axioma continuității). Pentru orice șir de evenimente $(A_n)_{n\in\mathbb{N}}$, $A_n \in \mathcal{K}$, $A_n \subseteq A_{n+1}$

$$\lim_{n\to\infty} I(A_n) = I(\lim_{n\to\infty} A_n) = I(\bigcup_{n\in\mathbb{N}} A_n) \tag{2.2.4}$$

A.6 (Axioma independenței). Dacă evenimentele $A, B \in \mathcal{K}$ sînt independente, atunci și numai atunci

$$I(A \cap B) = I(A) + I(B) \tag{2.2.5}$$

Să remarcăm că relațiile (2.2.3) și (2.2.5) pot fi formulate fără dificultate și în cazul unei reuniuni sau intersecții finite avînd

$$I\left(\bigcup_{k=1}^{n} A_k\right) = f\left(\sum_{k=1}^{n} f^{-1}(I(A_k))\right) \tag{2.2.6}$$

$$(\forall)\ A_k \in \mathcal{K},\ A_k \cap A_j = \emptyset,\ 1 \leqslant k \neq j \leqslant n$$

$$I\left(\bigcap_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} I(A_k) \Leftrightarrow A_k \in \mathcal{K},\ 1 \leqslant k \leqslant n \text{ sînt independente}$$

în totalitate. (2.2.7)

Ca urmare a axiomelor 2 și 3 deducem că

$$0 \leqslant I(A) \leqslant \infty \qquad (\forall)\ A \in \mathcal{K} \tag{2.2.8}$$

Axioma A.6 mai poate fi numită și a aditivității informației deoarece conform acesteia, informația ce provine din surse independente este aditivă. Aditivitatea este cauza principală a aspectului logaritmic al informației.

Propoziția 2.2.1. Pentru orice A și $B \in \mathcal{K}$ avem

$$I(A \cup B) \leqslant \min\{I(A), I(B)\} \leqslant \frac{1}{2}[I(A) + I(B))] \leqslant$$

$$\leqslant \max\{I(A), I(B)\} \leqslant I(A \cap B) \tag{2.2.9}$$

Demonstrație. Deoarece $A \cap B \subseteq A \subseteq A \cup B$ și $A \cap B \subseteq B \subseteq A \cup B$, relația (2.2.9) este consecință a axiomei 3.

Scanned with OKEN Scanner

Relaţia (2.2.9) devine egalitate peste tot dacă şi numai dacă $A = B$.

Se poate observa că primele trei inegalităţi ale relaţiei (2.2.9) pot fi deduse şi din axioma 4, ultima inegalitate fiind trivială căci $I(A \cap B) = \infty$, pentru $A \cap B = \emptyset$.

Remarcăm de asemenea că relaţia (2.2.9) poate fi extinsă sub forma

$$I\left(\bigcup_{k=1}^{n} A_k\right) \leqslant \min_{1 \leqslant k \leqslant n} \{I(A_k)\} \leqslant \frac{1}{n} \sum_{k=1}^{n} I(A_k) \leqslant$$

$$\leqslant \max_{1 \leqslant k \leqslant n} \{I(A_k)\} \leqslant I\left(\bigcap_{k=1}^{n} A_k\right) \qquad (2.2.10)$$

Axiomele de mai sus ne conduc la concluzia că între şansele de realizare a unui eveniment şi cantitatea de informaţie obţinută în consecinţă există o anumită legătură. Cu cît evenimentul are şanse mai mari de realizare, cu atît este mai mică informaţia pe care realizarea sa o aduce. Cum şansele de realizare a evenimentelor A sînt date de probabilitatea $P(A)$, putem spune că în mod natural există o funcţie $\varphi : [0, 1] \to [0, \infty)$, strict descrescătoare, astfel încît

$$I(A) = \varphi[P(A)] \qquad (2.2.11)$$

Deducem imediat

Propoziţia 2.2.2. Dacă $\varphi(x) = -\log x$, atunci relaţia (2.2.11) defineşte o măsură a cantităţii de informaţie în sensul axiomelor A.1 — A.6.

Să observăm că $\varphi(x) = \frac{1}{x} - x$ îndeplineşte condiţiile din (2.2.11) dar nu mai defineşte o măsură a informaţiei căci nu toate axiomele sînt îndeplinite. Dacă informaţia pe care o aduce realizarea unui eveniment A poate fi măsurată, se pune întrebarea: care este informaţia care se obţine în urma efectuării unui experiment ce poate avea n rezultate, deci care poate conduce la n evenimente elementare şi care, fiecare în parte, nu furnizează aceeaşi cantitate de informaţie? Răspunsul, utilizînd cantitatea de informaţie individuală, corespunzătoare fiecărui eveniment elementar, a fost dat de Shannon prin introducerea conceptului de entropie, sau mai degrabă de entropie informaţională.

Scanned with OKEN Scanner

3. Entropia Shannon

3.1. Entropia unei variabile aleatoare discrete

Fie $X : \begin{pmatrix} x_i \\ p_i \end{pmatrix}$, $x_i \in \mathbb{R}$, $p_i = P(X = x_i)$, $0 \leqslant p_i \leqslant 1$, $1 \leqslant i \leqslant n$, $\sum_{i=1}^{n} p_i = 1$, o variabilă aleatoare discretă. Dacă realizarea evenimentului $\{X = x_i\}$ furnizează o cantitate de informație $I(\{X = x_i\})$, atunci în mod natural efectuarea experimentului ale cărui rezultate sînt evenimentele $\{X = x_i\}$, $1 \leqslant i \leqslant n$, deci a experimentului descris de variabila aleatoare X, ar putea furniza o cantitate medie de informație $H(X)$ care să satisfacă inegalitatea

$$\min_{1 \leqslant i \leqslant n} \{I(\{X = x_i\})\} \leqslant H(X) \leqslant \max_{1 \leqslant i \leqslant n} \{I(\{X = x_i\})\} \quad (3.1.1)$$

Definiția 3.1.1. Se numește entropie Shannon a variabilei aleatoare discrete X mărimea

$$H(X) = H_n(p_1, p_2, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log p_i \quad (3.1.2)$$

Să observăm că $H(X)$ este cantitatea medie de informație obținută prin efectuarea experimentului cu n rezultate posibile, descris de variabila aleatoare X. Dacă notăm cu $\mathfrak{X}$ mulțimea tuturor variabilelor aleatoare discrete definite pe Ω, atunci putem considera un operator

$$I : \mathfrak{X} \to \{I(X)/X \in \mathfrak{X};\ I(X) : \begin{pmatrix} I(x_i) \\ p_i \end{pmatrix}, I(x_i) = I(\{X = x_i\}),$$
$$1 \leqslant i \leqslant n\} \quad (3.1.3)$$

pe care îl numim operatorul cantitate de informație și ca urmare, putem scrie că $H(X)$ este valoarea medie a variabilei aleatoare discrete $I(X)$, adică,

$$H(X) = M[I(X)] \quad (3.1.4)$$

Relația (3.1.2) definind o măsură a cantității medii de informație furnizată de un experiment cu n rezultate posibile, s-au propus și unități de măsură corespunzătoare. Astfel, considerînd drept cel mai simplu experiment pe acela

Scanned with OKEN Scanner

cu două rezultate egal probabile, cantitatea de informație corespunzătoare este luată ca unitate de măsură a cantității de informație. Avem deci

$$1 \text{ u.i.} = H_2\left(\frac{1}{2}, \frac{1}{2}\right) = \log 2. \tag{3.1.5}$$

Shannon, considerînd logaritmul în baza doi, a propus în 1948 ca această unitate să se numească bit (de la denumirea engleză binary digit, ceea ce înseamnă cifră binară și se referă la 0 și 1 care alcătuiesc sistemul binar și cu care orice număr poate fi scris în baza doi).

S-au mai propus denumirile „hartley", în 1950, în cazul considerării logaritmilor zecimali sau „hut", în 1952, în cazul considerării logaritmilor naturali. Cum trecerea de la o bază a logaritmului la alta nu face decît o multiplicare cu o constantă a cantității de informație existente, considerarea uneia sau alteia din denumiri este neesențială. În cele ce urmează vom considera peste tot baza doi, fără a mai menționa acest fapt, bază utilizată peste tot mai ales în limbajul calculatoarelor. Cînd vom considera și o altă bază vom specifica acest lucru.

Propoziția 3.1.1. Entropia Shannon a unei variabile aleatoare $H(X)$ are următoarele proprietăți:

(P.1) este definită pentru orice X

(P.2) este simetrică în raport cu probabilitățile $\{p_i\}_{i=1}^n$

(P.3) $H(X) \geqslant 0$, cu egalitate dacă și numai dacă X este constantă

(P.4) $H(X) \leqslant H_n\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) = \log n$, cu egalitate dacă și numai dacă X este repartizată uniform

(P.5) $H_{n+1}(p_1, \ldots, p_n, 0) = H_n(p_1, \ldots, p_n)$

(P.6) este invariantă în raport cu orice transformare biunivocă φ a lui X, adică $H[\varphi(X)] = H(X)$,

Demonstrație. Din definiția lui $H(X)$ și cu convenția $0 \log 0 = 0$, (P.1) — (P.3) și (P.5) sînt evidente. (P.4) rezultă ca o problemă simplă de extremum cu legături pentru funcția $H_n(p_1, \ldots, p_n)$ cu condiția $p_1 + p_2 + \ldots + p_n = 1$. (P.6) rezultă din faptul că dacă φ este biunivocă X și $\varphi(X)$ au aceeași repartiție.

Să remarcăm că din (P.3) și (P.4) rezultă

$$0 \leqslant H(X) \leqslant \log n \tag{3.1.6}$$

Scanned with OKEN Scanner

valoarea maximă a entropiei Shannon fiind tocmai expresia cantității de informație propusă de Hartley în 1927. Ea nu poate fi acceptată deoarece acordă aceeași cantitate de informație fiecărui eveniment posibil al unui experiment oarecare, indiferent de șansele de realizare ale acestuia, fapt care în general nu corespunde realității.

Să remarcăm, pentru rigoarea prezentării, că în relația (3.1.2) nu se face nici o precizare privind valorile pe care le poate lua n.

Se impune deci prezentarea unei forme mai generale a relației (3.1.2) ca fiind

$$H(X) = -\sum_{n=1}^{\infty} p_n \log p_n \tag{3.1.7}$$

expresie ce ar corespunde entropiei unui experiment cu o mulțime numărabilă de rezultate posibile. Dacă notăm cu $M \subseteq \mathbf{N}$ mulțimea rezultatelor posibile ale unui experiment oarecare, atunci relațiile (3.1.2) și (3.1.7) pot fi regăsite în relația

$$H(X) = -\sum_{n \in M} p_n \log p_n \tag{3.1.8}$$

după cum M este finită sau numărabilă. Adeseori relația (3.1.8) mai poate fi scrisă sub forma

$$H(X) = -\sum_{x} p(x) \log p(x) \tag{3.1.9}$$

înțelegînd că suma se face după toate valorile variabilei aleatoare X, cu condiția ca seria numerică cu termeni pozitivi (3.1.9) sau (3.1.7) să fie convergentă. Propoziția 3.1.1 rămîne valabilă în continuare și pentru entropia (3.1.8), proprietățile (P.4) și (P.5) fiind acum triviale.

În baza relației (3.1.8) putem deduce fără dificultate următoarele :

Propoziția 3.1.2. Dacă X este repartizată binomial, adică $p(x) = C_n^x p^x q^{n-x}$, $x = 0, 1, \ldots, n$; $p, q \in (0, 1)$, $p + q = 1$, atunci

$$H(X) = -n[p \log p + q \log q] - \sum_{x=0}^{n} C_n^x p^x q^{n-x} \log C_n^x \tag{3.1.10}$$

Scanned with OKEN Scanner

Propoziția 3.1.3. Dacă X este repartizată hipergeometric adică $p(x) = \dfrac{C_{N_1}^x C_{N_2}^{n-x}}{C_N^n}$, $x = 0, 1, \ldots, n$; $n \leqslant \min\{N_1, N_2\}$, $N = N_1 + N_2$, atunci

$$H(X) = \log C_N^n - \frac{1}{C_N^n} \sum_{x=0}^{n} C_{N_1}^x C_{N_2}^{n-x} \log C_{N_1}^x C_{N_2}^{n-x} \quad (3.1.11)$$

Propoziția 3.1.4. Dacă X este repartizată Poisson, adică $p(x) = \dfrac{\lambda^x}{x!} e^{-\lambda}$, $x \in \mathbf{N}$, $\lambda > 0$, atunci

$$H(X) = \lambda \log \frac{\lambda}{e} + e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \log x! \quad (3.1.12)$$

Dacă ținem cont că $x - 1 \leqslant \log_2 x! \leqslant \dfrac{x(x-1)}{2}$, deducem imediat din (3.1.12) că

$$\log \frac{e^{2\lambda - 1}}{\lambda^\lambda} \leqslant H(X) \leqslant \log \frac{e^{\frac{\lambda^2}{2} + 1}}{\lambda^\lambda} \quad (3.1.13)$$

ceea ce confirmă odată în plus convergența seriei (3.1.12) și deci a lui $H(X)$.

Propoziția 3.1.5. Dacă X este repartizată geometric, adică $p(x) = pq^x$, $x \in \mathbf{N}$, $0 < p, q < 1$, $p + q = 1$, atunci

$$H(X) = -\frac{1}{p} [p \log p + q \log q] \quad (3.1.14)$$

Se observă că exceptînd repartițiile uniformă și geometrică, pentru care se poate da o formulă generală a entropiei, în celelalte cazuri aceasta se calculează de fiecare dată, ca pentru orice variabilă aleatoare arbitrară.

3.2. Entropia unei variabile continue

3.2.1. Entropia Boltzmann-Shannon-Wiener

Fie X o variabilă aleatoare continuă, cu densitatea de repartiție $p(x) \geqslant 0$, $\int_{\mathbf{R}} p(x)\, dx = 1$, descriind un anumit ex-

Scanned with OKEN Scanner

periment aleator continuu. Pornind de la definiția entropiei corespunzătoare unui experiment aleator discret, N. Wiener și C. E. Shannon au considerat, prin analogie, drept măsură a cantității de informație furnizată de un experiment aleator continuu, expresia

$$H(X) = -\int_{\mathbf{R}} p(x) \log p(x)\, dx \tag{3.2.1}$$

care este analogă celebrei funcții H a lui L. Boltzmann ce definea „negentropia" (entropia cu semn schimbat) unui gaz în teoria cinetică moleculară și a cărei expresie era

$$H = \int_{\mathbf{R}^3} f(x, y, z) \log f(x, y, z)\, dx\, dy\, dz \tag{3.2.2}$$

unde $f(x, y, z)$ este densitatea de probabilitate a vitezei de componente (x, y, z) a unei molecule de gaz, motiv pentru care numim $H(X)$ dată de (3.2.1) entropia Boltzmann-Wiener-Shannon.

Dacă între relațiile (3.1.2) și (3.2.1) există o analogie ca urmare a semnificației lor de cantități de informație corespunzătoare unui experiment aleator discret și respectiv continuu, proprietățile entropiei nu pot fi transpuse prin analogie din cazul discret în cazul continuu.

Propoziția 3.2.1. Entropia unei variabile aleatoare continue

a) nu este o limită a entropiei unei variabile aleatoare discrete

b) nu este totdeauna nenegativă

c) nu este invariantă față de o transformare biunivocă a variabilei.

Demonstrație. a) Într-adevăr, să presupunem că X este repartizată uniform continuu pe $[a, b]$ adică

$$p(x) = \begin{cases} \dfrac{1}{b-a} & x \in [a, b] \\ 0 & x \in [a, b] \end{cases} \tag{3.2.3}$$

Deducem din (3.2.1) și (3.2.3) că $H(X) = \log(b - a)$. Dacă vom considera acum o diviziune $\{A_k\}_{k=1}^n$ a lui $[a, b]$,

Scanned with OKEN Scanner

$A_k \cap A_l = \emptyset$, $1 \leqslant k \neq l \leqslant n$, $\bigcup_{k=1}^{n} A_k = [a, b]$, atunci putem considera variabila aleatoare discretă Y avînd repartiția

$$p = P(A_k) = P(y \in A_k) = \frac{1}{n}, \quad 1 \leqslant k \leqslant n \qquad (3.2.4)$$

corespunzător căreia $H(Y) = \log n$ în baza relației (3.1.2). Considerînd acum partiția lui $[a, b]$ din ce în ce mai fină, făcînd deci pe n să tindă la infinit, deducem că $\log (b - a) \neq \neq \lim_{n \to \infty} \log n = \infty$ ceea ce justifică afirmația a). Tot pentru repartiția uniformă, să observăm că $H(X) = \log(b - a) \leqslant \leqslant 0$ pentru orice interval $[a, b]$ de lungime subunitară, adică $b - a \leqslant 1$. Mai mult chiar, dacă $b - a = 1$, $H(X) = = 0$, ceea ce în cazul discret este adevărat dacă și numai dacă variabila ar fi constantă cu probabilitatea unu. Afirmația b) fiind justificată, pentru afirmația c) să reluăm tot repartiția uniformă. Deducem imediat că dacă X este repartizată uniform pe $[a, b]$. atunci variabila $Y = \alpha X$ este de asemenea uniform repartizată, cu

$$p(y) = \begin{cases} \dfrac{1}{\alpha(b - a)} & y \in [\alpha a, \ \alpha b] \\ 0 & y \in [\alpha a, \ \alpha b] \end{cases} \qquad (3.2.5)$$

și ca urmare $H(Y) = \log \alpha(b - a) \neq \log (b - a) = H(X)$. De altfel rezultatele de mai sus, cu deosebire b) și c) pot fi constatate ușor și pentru alte tipuri de repartiții. Pozitivitatea entropiei poate fi asigurată luînd $|-\log p(x)|$.

În baza relației (3.2.1), prin calcul direct, putem deduce expresia entropiei pentru repartițiile continue clasice unidimensionale.

Propoziția 3.2.2. Dacă X este normal repartizată (sau Gauss-Laplace), adică

$$p(x, m, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \ x \in \mathbf{R}, \ m \in \mathbf{R}, \ \sigma > 0 \qquad (3.2.6)$$

atunci

$$H(X) = \log \sigma \sqrt{2\pi e} \qquad (3.2.7)$$

Se observă că nedeterminarea corespunzătoare unui experiment aleator de repartiție normală este cu atît mai mare cu cît este mai mare abaterea medie pătratică σ a acestuia.

Scanned with OKEN Scanner

Propoziția 3.2.3. Dacă X este repartizată Laplace, adică

$$p(x, \lambda, \mu) = \frac{1}{2\lambda} e^{-\frac{|x-\mu|}{\lambda}}, \quad x \in \mathbf{R}, \quad \mu \in \mathbf{R}, \quad \lambda > 0 \quad (3.2.8)$$

atunci

$$H(X) = \log 2\lambda e \quad (3.2.9)$$

Propoziția 3.2.4. Dacă X este repartizată exponențial, adică

$$p(x, \lambda) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}}, \quad x \geqslant 0, \quad \lambda > 0 \quad (3.2.10)$$

atunci

$$H(X) = \log \lambda e \quad (3.2.11)$$

rezultat ce poate fi dedus și din relația (3.2.9).

Propoziția 3.2.5. Dacă X este repartizată gama, adică

$$p(x, \alpha) = \frac{x^{\alpha-1} e^{-x}}{\Gamma(\alpha)}, \quad x \geqslant 0, \quad \alpha > 0, \quad \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \quad (3.2.12)$$

atunci

$$H(X) = \log \left[\Gamma(\alpha) \exp \left\{ \frac{\alpha^2 + (\alpha - 1)(1 + \alpha\gamma)}{\alpha} - \right. \right.$$

$$\left. \left. - \alpha(\alpha - 1) \sum_{n=1}^{\infty} \frac{1}{n(n + \alpha)} \right\} \right] \quad (3.2.13)$$

unde $\gamma = \lim_{n\to\infty} \left[1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} - \ln n \right]$ este constanta lui Euler $\approx 0{,}577215$.

Demonstrație. Într-adevăr, avem

$$H(X) = \left[\ln \Gamma(\alpha) - \frac{\alpha - 1}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-x} \ln x \, dx + \alpha \right] \log e =$$

$$= \log \Gamma(\alpha) e^{\alpha - (\alpha-1)\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}}$$

Scanned with OKEN Scanner

unde $\Gamma'(\alpha)$ este derivata funcției $\Gamma(\alpha)$ din (3.2.12). Ținînd cont de formula lui Weierstrass

$$\frac{1}{\Gamma(\alpha)} = \alpha\, e^{\gamma\alpha} \prod_{n=1}^{\infty} \left(1 + \frac{\alpha}{n}\right) e^{-\frac{\alpha}{n}} \tag{3.2.14}$$

obținem

$$-\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} = \frac{d}{d\alpha}\left[\ln \frac{1}{\Gamma(\alpha)}\right] = \frac{1}{\alpha} + \gamma - \alpha \sum_{n=1}^{\infty} \frac{1}{n(n+\alpha)}$$

raport care există pentru orice $\alpha > 0$, seria $\sum_{n=1}^{\infty} \frac{1}{n(n+\alpha)}$ fiind convergentă și ca urmare rezultă relația (3.2.13).

Să remarcăm că pentru $\alpha = 1$ și $\lambda = 1$, relațiile (3.2.13) și (3.2.11) coincid, fiind vorba de aceeași repartiție particulară.

Propoziția 3.2.6. Dacă X este repartizată Beta, adică

$$p(x, \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, \quad x \in [0, 1], \quad \alpha > 0, \quad \beta > 0,$$

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\, dx \tag{3.2.15}$$

atunci

$$H(X) = \log\left[B(\alpha, \beta)\exp\left\{\sum_{n=0}^{\infty} \frac{1}{n+\alpha+\beta}\left(\frac{(\alpha-1)\beta}{n+\alpha} + \frac{\alpha(\beta-1)}{n+\beta}\right)\right\}\right] \tag{3.2.16}$$

Demonstrație. Într-adevăr, avem

$$H(X) = \left[\ln B(\alpha, \beta) - \frac{\alpha-1}{B(\alpha, \beta)} \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} \ln x\, dx - \right.$$

$$\left. - \frac{\beta-1}{B(\alpha, \beta)} \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} \ln(1-x)\, dx\right] \log e =$$

Scanned with OKEN Scanner

$$= \log\left[B(\alpha, \beta) \exp\left\{ (1-\alpha) \frac{\partial}{\partial \alpha} \ln B(\alpha, \beta) + (1-\beta) \frac{\partial}{\partial \beta} \ln B(\alpha, \beta) \right\} \right]$$

Ca urmare a relației

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\ \Gamma(\beta)}{\Gamma(\alpha + \beta)} \qquad (3.2.17)$$

obținem

$$\frac{\partial}{\partial \alpha} [\ln B(\alpha, \beta)] = \frac{\partial}{\partial \alpha} \left[\ln \frac{\Gamma(\alpha)}{\Gamma(\alpha + \beta)} \right] = -$$

$$- \beta \sum_{n=0}^{\infty} \frac{1}{(n + \alpha)(n + \alpha + \beta)}$$

$$\frac{\partial}{\partial \beta} [\ln\ B(\alpha, \beta)] = \frac{\partial}{\partial \beta} \left[\ln \frac{\Gamma(\beta)}{\Gamma(\alpha + \beta)} \right] = -$$

$$- \alpha \sum_{n=0}^{\infty} \frac{1}{(n + \beta)(n + \alpha + \beta)}$$

și în consecință rezultă relația (3.2.16). Expresiile de mai sus și ca urmare relația (3.2.16), pot fi obținute și prin derivarea funcției

$$B(\alpha, \beta) = \frac{\alpha + \beta}{\alpha \beta} \prod_{n=1}^{\infty} \frac{(n + \alpha + \beta)}{(n + \alpha)(n + \beta)} \qquad (3.2.18)$$

a cărei exprimare este consecință a relațiilor (3.2.14) și (3.2.17).

Propoziția 3.2.7. Dacă X este repartizată $H(m, \sigma)$, adică

$$p(x, m, \sigma) = \frac{x^{\frac{m}{2}-1} e^{-\frac{x}{2\sigma^2}}}{2^{\frac{m}{2}} \sigma^m \Gamma\left(\frac{m}{2}\right)}, \quad x \geqslant 0,\ \sigma > 0,\ m \in \mathbf{N} \qquad (3.2.19)$$

atunci

$$H(X) = \log \left[2\sigma^2 \Gamma\left(\frac{m}{2}\right) \exp\left\{ \frac{m^2 + (m-2)(2 + m\gamma)}{2m} - \right.\right.$$

$$\left.\left. - \frac{m(m-2)}{2} \sum_{n=1}^{\infty} \frac{1}{n(2n + m)} \right\} \right] \qquad (3.2.20)$$

unde γ și $\Gamma\left(\frac{m}{2}\right)$ sînt date mai sus, în propoziția 3.2.5

Scanned with OKEN Scanner

Demonstrație. Punînd formal $\frac{m}{2} = \alpha$ și ținînd cont de derivata funcției ln $\Gamma(\alpha)$ dată mai sus, rezultă (3.2.20).

Să remarcăm că pentru $m = 2\alpha$ și $2\sigma^2 = 1$, relațiile (3.2.20) și (3.2.21) coincid.

Propoziția 3.2.8. Dacă X este repartizată Student, adică

$$p(x) = \frac{1}{\sqrt{m}\, B\left(\frac{1}{2}, \frac{m}{2}\right)} \left(1 + \frac{x^2}{m}\right)^{-\frac{m+1}{2}}, \quad x \in \mathbf{R}, \ m \in \mathbf{N} \tag{3.2.21}$$

atunci

$$H(X) = \log\left[\sqrt{m}\, B\left(\frac{1}{2}, \frac{m}{2}\right) \exp\left\{(m + 1) \sum_{n=0}^{\infty} \frac{1}{(2n + m)(2n + m + 1)}\right\}\right] \tag{3.2.22}$$

Demonstrație. Într-adevăr, avem

$$H(X) = \left[\ln \sqrt{m}\, B\left(\frac{1}{2}, \frac{m}{2}\right) + \right.$$

$$\left. + \frac{m+1}{\sqrt{m}\, B\left(\frac{1}{2}, \frac{m}{2}\right)} \int_0^{\infty} \left(1 + \frac{x^2}{m}\right)^{-\frac{m+1}{2}} \ln\left(1 + \frac{x^2}{m}\right) dx \right] \log e$$

Făcînd schimbarea de variabilă

$$x^2 = \frac{mt}{1 - t}, \quad t \in [0, 1)$$

și ținînd cont de derivatele funcției ln $B(\alpha, \beta)$ date mai sus obținem relația (3.2.22).

Un caz particular al relației (3.2.22) se obține pentru $m = 1$, cînd avem $H(X) = \log 4\pi$, adică entropia reparti-

Scanned with OKEN Scanner

ției Cauchy, sau a repartiției Student cu un grad de libertate, a cărei densitate este

$$p(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbf{R} \tag{3.2.23}$$

Propoziția 3.2.9. Dacă X este repartizată Snedecor, adică

$$p(x, n_1, n_2) = \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} \frac{x^{\frac{n_1}{2}-1}\left(1+\frac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)}, \; x \geqslant 0, \; n_1, n_2 \in \mathbf{N} \tag{3.2.24}$$

atunci

$$H(X) = \log\left[\frac{n_2}{n_1} B\left(\frac{n_1}{2}, \frac{n_2}{2}\right) \exp\left\{\sum_{n=0}^{\infty} \frac{1}{(2n+n_1+n_2)}\left(\frac{n_2(n_1-2)}{2n+n_1} + \frac{n_1(n_2+2)}{2n+n_2}\right)\right\}\right] \tag{3.2.25}$$

Demonstrație, Într-adevăr,

$$H(X) = \left[\ln\left(\frac{n_2}{n_1}\right)^{\frac{n_1}{2}} B\left(\frac{n_1}{2}, \frac{n_2}{2}\right) - \frac{n_1-2}{2} \cdot \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} B^{-1}\left(\frac{n_1}{2}, \frac{n_2}{2}\right) \times \right.$$

$$\times \int_0^{\infty} x^{\frac{n_1}{2}-1}\left(1+\frac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}} \ln x \, dx +$$

$$+ \frac{n_1+n_2}{2}\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} B^{-1}\left(\frac{n_1}{2}, \frac{n_2}{2}\right) \times$$

$$\left. \times \int_0^{\infty} x^{\frac{n_1}{2}-1}\left(1+\frac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}} \ln\left(1+\frac{n_1}{n_2}x\right) dx\right] \cdot \log e.$$

Scanned with OKEN Scanner

Efectuînd schimbarea de variabilă $x = \frac{n_2 t}{n_1(1-t)}$, $t \in [0, 1)$, şi ţinînd cont de derivatele funcţiei ln $B(\alpha, \beta)$ date mai sus, deducem relaţia (3.2.25).

Propoziţia 3.2.10. Dacă X este repartizată Fisher, adică

$$p(x, n_1, n_2) = 2\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} \frac{e^{n_1 x}\left(1 + \frac{n_1}{n_2} e^{2x}\right)^{-\frac{n_1+n_2}{2}}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)}, \quad x \in \mathbf{R},$$

$$n_1, n_2 \in \mathbf{N} \tag{3.2.26}$$

atunci

$$H(X) = \log\left[\frac{1}{2} B\left(\frac{n_1}{2}, \frac{n_2}{2}\right) \exp\left\{\sum_{n=0}^{\infty} \frac{1}{2n + n_1 + n_2}\left(\frac{2n_1 n_2}{2n + n_1} + \right.\right.\right.$$

$$\left.\left.\left. + \frac{n_1(n_2 - n_1)}{2n + n_2}\right)\right]\right\} \tag{3.2.27}$$

Demonstraţie. Într-adevăr, scriind $H(X)$ pentru (3.2.26) şi efectuînd schimbarea de variabilă

$$e^x = \sqrt{\frac{n_2}{n_1}} \frac{t}{1-t}, \quad t \in [0, 1)$$

şi ţinînd iarăşi cont de derivatele funcţiei $\ln B(\alpha, \beta)$, deducem relaţia (3.2.27).

3.2.2. Entropia ca variaţie a informaţiei

Pornind de la definiţia entropiei Shannon şi avînd în vedere că trecerea de la cazul discret la cazul continuu nu înseamnă o trecere la limită de la discret la continuu şi în

Scanned with OKEN Scanner

expresia entropiei, S. Kullback și R. A. Leibler au avut în 1951 ideea ca, în loc de a vorbi de entropia atașată unei anumite probabilităţi izolate, să se compare entropiile în interiorul unei clase de măsuri. Ideea și rezultatele de atunci au fost dezvoltate ulterior în alte lucrări ale lor sau ale altor matematicieni între care J. Kampé de Fériet (1963). Fie P_1 și P_2 două măsuri ale probabilității, sau două probabilități, definite pe cîmpul de evenimente $\{\Omega, \mathcal{K}\}$. Să presupunem că P_2 socotită ca probabilitate inițială, este înlocuită cu P_1. Atunci variaţia informaţiei furnizată de evenimentul $A \in \mathcal{K}$ va fi

$$P_1(A)[(-\log P_2(A)) - (-\log P_1(A))] = P_1(A)\log \frac{P_1(A)}{P_2(A)} \tag{3.2.28}$$

și ca urmare variaţia totală de informaţie corespunzătoare partiției finite $\{A_k\}_{k=1}^n$ a lui Ω cînd trecem de la repartiţia P_2 la repartiția P_1 va fi notată cu $I(P_1/P_2)$ și va avea expresia

$$I(P_1/P_2) = \sum_{k=1}^{n} P_1(A_k) \log \frac{P_1(A_k)}{P_2(A_k)} \tag{3.2.29}$$

care poate fi finită sau infinită, ultimul caz avînd loc dacă există un $A_{k_0} \in \mathcal{K}$ astfel că $P_1(A_{k_0}) \neq 0$ și $P_2(A_{k_0}) = 0$, ceea ce se întîmplă dacă P_1 și P_2 sînt măsuri reciproc singulare (adică dacă $P_i(A) = 0 \nrightarrow P_j(A) = 0$, $1 \leqslant i \neq j \leqslant 2$). Dacă P_1 este absolut continuă în raport cu P_2, ceea ce înseamnă că $P_2(A) = 0 \Rightarrow P_1(A) = 0$ și scriem $P_1 \ll P_2$, atunci $H(P_1/P_2) < \infty$.

Fie acum μ_1 și μ_2 două măsuri ale probabilităţii definite pe $\mathcal{K}$, astfel că μ_1 este absolut continuă în raport cu μ_2, adică $\mu_1 \ll \mu_2$. Conform cunoscutei teoreme a lui Radon-Nikodym, există o funcţie $\varphi(x)$ nenegativă definită pe Ω astfel încît pentru orice $E \in \mathcal{K}$ să avem

$$\mu_1(E) = \int_E \varphi(x)\, d\mu_2(x), \tag{3.2.30}$$

iar în cazul particular $E = \Omega$

$$1 = \int_{\Omega} \varphi(x)\, d\mu_2(x). \tag{3.2.31}$$

Asta înseamnă de fapt că $\varphi(x)$ este μ_2 — integrabilă pe E.

Definiția 3.2.1. Se numește variație a informației cînd se trece de la măsura probabilității μ_2 la măsura probabilității μ_1, absolut continuă în raport cu μ_2, integrala

$$I(\mu_1/\mu_2) = \int_{\Omega} \varphi(x) \log \varphi(x)\, d\mu_2(x) = \int_{\Omega} \log \varphi(x)\, d\mu_1(x) \tag{3.2.32}$$

Vom da acum fără demonstrații, care sînt bine prezentate în lucrările lui S. Kullback (1959), J. Kampé de Fériet (1963) și S. Guiașu (1968, 1977), cîteva rezultate asupra măsurii variației informației $I(\mu_1/\mu_2)$.

Propoziția 3.2.11. Fie $\mu_1 \ll \mu_2$. Pentru orice $E \in \mathcal{K}$ avem

$$\mu_1(E) \log \frac{\mu_1(E)}{\mu_2(E)} \leqslant I(\mu_1/\mu_2) \leqslant$$

$$\leqslant \mu_1(E) \log \left[\frac{1}{\mu_1(E)} \int_{\Omega} \varphi^2(x)\, d\mu_2(x) \right] \tag{3.2.33}$$

cu prima egalitate dacă și numai dacă $\varphi(x) = \dfrac{\mu_1(E)}{\mu_2(E)}$ μ_1 - aproape peste tot pe E și cu a doua egalitate numai dacă $\varphi(x) =$ constant μ_1 — aproape peste tot pe E.

Ca o consecință a propoziției 3.2.11, punînd $E = \Omega$ deducem

Propoziția 3.2.12. Avem

$$0 \leqslant I(\mu_1/\mu_2) \leqslant \log \left[\int_{\Omega} \varphi^2(x)\, d\mu_2(x)\, dx \right] = \log \left[\int_{\Omega} \varphi(x)\, d\mu_1(x) \right] \tag{3.2.34}$$

Scanned with OKEN Scanner

Conform relației (3.2.34) rezultă că dacă $\varphi^2(x)$ este μ_2-integrabilă, sau, dacă $\varphi(x)$ este μ_1-integrabilă pe Ω, atunci variația de informație $I(\mu_1/\mu_2)$ este finită.

Să considerăm acum o partiție finită $\mathcal{A} = \{A_k\}_{k=1}^n$ a lui Ω. Ca urmare a propoziției 3.2.11 avem

$$\sum_{k=1}^{n} \mu_1(A_k)) \log \frac{\mu_1(A_k)}{\mu_2(A_k)} \leqslant I(\mu_1/\mu_2) \qquad (3.2.35)$$

ceea ce implică inegalitatea

$$\sup_{\mathcal{A}} \sum_{k=1}^{n} \mu_1(A_k) \log \frac{\mu_1(A_k)}{\mu_2(A_k)} \leqslant I(\mu_1/\mu_2) \qquad (3.2.36)$$

înțelegînd că supremum este luat după toate partițiile finite măsurabile $\mathcal{A}$ ale lui Ω. Următoarea propoziție, a cărei demonstrație nu o dăm aici, ne asigură că relația (3.2.36) este de fapt o egalitate.

Propoziția 3.2.13. Avem

$$\sup_{\Omega} \sum_{k=1}^{n} \mu_1(A_k) \log \frac{\mu_1(A_k)}{\mu_2(A_k)} = I(\mu_1/\mu_2) \qquad (3.2.37)$$

unde supremum este luat după toate partițiile finite măsurabile ale mulțimii Ω.

Ca urmare a celor de mai sus, să considerăm cazul particular în care Ω conține n evenimente elementare, avînd

μ_1 definită de repartiția $p_k \geqslant 0$, $\sum_{k=1}^{n} p_k = 1$

μ_2 definită de repartiția $q_k \geqslant 0$, $\sum_{k=1}^{n} q_k = 1$.

Dacă $\mu_1 \ll \mu_2$ atunci putem lua $\varphi(x_k) = \frac{p_k}{q_k}$ ca derivata Radon-Nicodym a lui μ_1 în raport cu μ_2 și ca urmare

$$I(\mu_1/\mu_2) = \sum_{k=1}^{n} p_k \log p_k - \sum_{k=1}^{n} p_k \log q_k. \qquad (3.2.38)$$

Dacă $q_k = \frac{1}{n}$, $1 \leqslant k \leqslant n$, atunci din (3.2.38) deducem că

$$H_n(p_1, \ldots, p_n) = \log n - I(\mu_1/\mu_2) \qquad (3.2.39)$$

Scanned with OKEN Scanner

și în consecință observăm că entropia Shannon apare ca fiind, cu excepția unei constante aditive, variația cantității de informație cînd trecem de la repartiția probabilistică uniformă la repartiția dată de numerele $p_k \geqslant 0$, $\sum_{k=1}^{n} p_k = 1$ (orice astfel de distribuție probabilistă este absolut continuă în raport cu distribuția uniformă discretă).

Să considerăm acum cazul particular $\Omega = [a, b]$, avînd

μ_1 definită de repartiția $p(x) \geqslant 0, \int_a^b p(x)\,dx = 1$

μ_2 definită de repartiția $q(x) \geqslant 0, \int_a^b q(x)\,dx = 1$

Dacă $\mu_1 \ll \mu_2$ atunci derivata Radon-Nicodym a lui μ_1 în raport cu μ_2 este $\varphi(x) = \frac{p(x)}{q(x)}$ și ca urmare

$$I(\mu_1/\mu_2) = \int_a^b p(x) \log \frac{p(x)}{q(x)}\,dx \qquad (3.2.40)$$

relație din care, în cazul particular în care $q(x)$ este densitatea de repartiție uniformă continuă pe $[a, b]$, deci $q(x) = \frac{1}{b-a}$ pentru $x \in [a, b]$, deducem că

$$H_p(x) = \log(b-a) - I(\mu_1/\mu_2) \qquad (3.2.41)$$

ceea ce ne permite să interpretăm entropia Shannon continuă ca fiind, cu excepția unei constante aditive, variația cantității de informație cînd se trece de la repartiția uniformă pe $[a, b]$ la o nouă repartiție de densitate $p(x)$ (orice astfel de repartiție fiind absolut continuă în raport cu distribuția uniformă continuă pe $[a, b]$).

Ca urmare a celor de mai sus, rezultă că atît în cazul discret cît și în cazul continuu, entropia Shannon apare ca fiind variația cantității de informație cînd se trece de la

Scanned with OKEN Scanner

repartiția inițială definită de repartiția uniformă la o nouă repartiție. Acest lucru ne permite să utilizăm în cazul continuu entropia Boltzmann (care l-a inspirat pe Shannon) la fel de bine ca și entropia Shannon din cazul discret.

3.3. Entropii ale unui vector aleator discret

Să presupunem că se efectuează simultan două sau mai multe experimente, fiecare avînd un anumit număr de rezultate posibile, realizarea unuia sau unora din rezultate putînd fi condiționată sau nu de realizarea altora dintre ele. Dacă în cazul unui singur experiment am asociat acestuia o variabilă aleatoare ale cărei valori x_i reprezintă rezultatele posibile ale experimentului, care apar cu anumite probabilități $p_i = p(x_i)$, atunci, printr-o simplă generalizare putem asocia unui grup de experimente ce se efectuează simultan un vector aleator, ale cărui componente sînt variabile aleatoare asociate fiecăruia din experimentele grupului. Ori de cîte ori vom vorbi despre un vector aleator ne vom gîndi la un grup de experimente, sau la un experiment multidimensional, sau la un experiment vectorial. După cum grupul conține două, trei sau n experimente, putem vorbi de experimente bidimensionale, tridimensionale sau n-dimensionale, descrise de vectori aleatori bidimensionali, tridimensionali sau n-dimensionali.

Un anumit grup de experimente poate fi gîndit și ca o reuniune de alte grupuri de experimente distincte sau nu, ceea ce poate conduce și la a considera familii de grupuri de experimente și prin urmare, familli de vectori aleatori, pentru care noțiunea de entropie poate fi introdusă prin analogie cu experimentele unidimensionale, pornind de la relația (3.1.2) și cu precădere de la relația (3.1.9), pe care, pentru simplificarea scrierii, o vom folosi mai des.

3.3.1. Cazul bidimensional

Fie $X = (X_1, X_2)$ un vector aleator bidimensional discret, scris detaliat sub forma $X : \begin{pmatrix} (x_1, x_2) \\ p(x_1, x_2) \end{pmatrix}$, $0 \leqslant p(x_1, x_2) \leqslant 1$, $\sum_{x_1} \sum_{x_2} p(x_1, x_2) = 1$, $x_1 \in A_1$, $x_2 \in A_2$, atît A_1 cît și A_2 fiind finite sau cel mult numărabile, arbitrare.

Scanned with OKEN Scanner

Definiția 3.3.1. Se numește entropie Shanpon directă a vectorului aleator bidimensional $X = (X_1 X_2)$, sau încă entropie comună, expresia

$$H(X) = H(X_1, X_2) = -\sum_{x_1} \sum_{x_2} p(x_1, x_2) \log p(x_1, x_2) \tag{3.3.1}$$

Dacă notăm cu $x = (x_1, x_2)$, atunci relația (3.3.1) este tocmai relația (3.1.9). Deoarece repartițiile componentelor X_1 și X_2 ale vectorului aleator $X = (X_1, X_2)$ sînt numite repartiții individuale, vom spune prin analogie că $H(X_1)$ și $H(X_2)$ sînt entropiile individuale ale vectorului aleator $X = (X_1, X_2)$. Ca urmare a faptului că avem un experiment bidimensional, realizarea unuia dintre experimentele individuale poate fi influențată de realizarea celuilalt sau poate influența realizarea celuilalt și în consecință, cantitatea de informație obținută prin realizarea unuia dintre ele nu este aceeași în cazul în care cele două experimente sînt sau nu independente. Sîntem conduși la a defini și entropia condiționată pornind de la faptul că $H(X_1, X_2)$ reprezintă cantitatea medie de informație furnizată de realizarea cuplului de experimente descrise de vectorul aleator (X_1, X_2). Să presupunem că realizarea evenimentului A_1 depinde de realizarea evenimentului B, realizare care se efectuează cu probabilitatea condiționată $P(A_1/B)$ și care furnizează cantitatea de informație $I(A_1/B) = -\log P(A_1/B)$. Dacă $A_1, \ldots, A_n$ reprezintă mulțimea rezultatelor posibile ale unui experiment, fiecare dintre ele fiind condiționat în realizare de B, atunci apare ca naturală considerarea cantității medii de informație, furnizată de fiecare eveniment condiționat.

Definiția 3.3.2. Se numește entropie a variabilei aleatoare X_1 condiționată de evenimentul $\{X_2 = x_2\}$, mărimea

$$H(X_1/x_2) = -\sum_{x_1} p(x_1/x_2) \log p(x_1/x_2). \tag{3.3.2}$$

Evident că pentru fiecare x_2 avem o anumită expresie de tip (3.3.2) și prin urmare apare ca necesar să considerăm o valoare medie a acestora, căci realizarea experimentului descris de X_1 depinde de întregul experiment X_2 și nu numai de unul din rezultatele acestuia.

Definiția 3.3.3. Se numește entropie a variabilei aleatoare X_1 condiționată de variabila aleatoare X_2, mărimea

$$H(X_1/X_2) = \sum_{x_2} p(x_2) H(X_1/x_2) \tag{3.3.3}$$

Scanned with OKEN Scanner

Deoarece $p(x_1, x_2) = p(x_2)p(x_1/x_2)$, din relațiile (3.3.2) și (3.3.3) rezultă

$$H(X_1/X_2) = -\sum_{x_1}\sum_{x_2} p(x_1, x_2) \log p(x_1/x_2) \qquad (3.3.4)$$

Propoziția 3.3.1. Avem

$$H(X_1) + H(X_2/X_1) = H(X_1, X_2) = H(X_2) + H(X_1/X_2) \qquad (3.3.5)$$

Demonstrație. Din (3.3.4) rezultă $H(X_1/X_2) =$

$$= -\sum_{x_1}\sum_{x_2} p(x_1, x_2) \log \frac{p(x_1, x_2)}{p(x_2)} = -$$

$$-\sum_{x_1}\sum_{x_2} p(x_1, x_2) \log p(x_1, x_2) + \sum_{x_1}\sum_{x_2} p(x_1, x_2) \log p(x_2) =$$

$$= H(x_1, x_2) - H(x_2) \text{ și ca urmare} \qquad (3.3.5).$$

Definiția 3.3.3. Vom spune că X_1 și X_2 sînt entropic independente dacă $H(X_1/X_2) = H(X_1)$. Să remarcăm, din relația (3.3.5) și definiția 3.3.3, că $H(X_1/X_2) = H(X_1) \Leftrightarrow H(X_2/X_1) = H(X_2)$ și ca urmare, independența entropică poate fi dată doar cu una dintre egalități.

Definiția 3.3.4. Vom spune că X_1 și X_2 sînt probabilistic independente dacă pentru orice cuplu de valori (x_1, x_2) ale acestora avem $p(x_1, x_2) = p(x_1)p(x_2)$.

Propoziția 3.3.2. X_1 și X_2 sînt entropic independente dacă și numai dacă sînt probabilistic independente.

Demonstrație. Dacă X_1 și X_2 sînt entropic independente, atunci $\sum_{x_1}\sum_{x_2} p(x_1, x_2) \log \frac{p(x_1, x_2)}{p(x_1)p(x_2)} = 0$. Pe de altă parte, deoarece $\log x \geqslant 1 - \frac{1}{x}$ cu egalitate dacă și numai dacă $x = 1$, rezultă că

$$0 = \sum_{x_1}\sum_{x_2} p(x_1, x_2) \log \frac{p(x_1, x_2)}{p(x_1)\, p(x_2)} \geqslant$$

$$\geqslant \sum_{x_1}\sum_{x_2} p(x_1, x_2) \left[1 - \frac{p(x_1)\, p(x_2)}{p(x_1, x_2)}\right] = 0$$

cu egalitate dacă și numai dacă $\frac{p(x_1, x_2)}{p(x_1)\, p(x_2)} = 1$ pentru orice cuplu (x_1, x_2) deci dacă și numai dacă X_1 și X_2 sînt proba-

Scanned with OKEN Scanner

bilistic independente și afirmația este demonstrată. Ca urmare, vom spune doar independență, fără alte precizări.

Definiția 3.3.5. Vom spune că:

a) X_1 este complet dependentă de X_2 și vom scrie X_1 c.d. X_2 dacă $H(X_1/X_2) = 0$;

b) X_1 și X_2 sînt reciproc complet dependente și vom scrie X_1 r.c.d. X_2 dacă $H(X_1/X_2) = H(X_2/X_1) = 0$;

c) X_1 și X_2 sînt egal dependente una de alta dacă $H(X_1/X_2) = H(X_2/X_1) = 0$.

Să observăm că X_1 r.c.d.$X_2 \Leftrightarrow X_1$c.d.X_2 și X_2 c.d.X_1.

Definiția 3.3.6. Se spune că matricea $M = (m_{ij})$ este diagonală dacă $m_{ij} = 0$ pentru orice $i \neq j$ și că ea este cvasidiagonală dacă se poate obține dintr-o matrice diagonală prin schimbarea între ele a două sau mai multe linii sau a două sau mai multe coloane ale acesteia.

Definiția 3.3.7. Spunem că X_1 și X_2 sînt identic repartizate dacă matricea repartiției comune $p(x_1, x_2)$ este diagonală sau cvasidiagonală.

Definiția 3.3.8. Spunem că vectorul aleator $X = (X_1, X_2)$ este diagonal repartizat în raport cu X_1 dacă pentru fiecare x_1 există un singur x_2^0 astfel că

$$p(x_1, x_2) = \begin{cases} p(x_1) , & x_2 = x_2^0 \\ 0 , & x_2 \neq x_2^0. \end{cases} \tag{3.3.6}$$

Dacă X este diagonal repartizat în raport cu X_1 și cu X_2 spunem că el este diagonal repartizat în totalitate sau că X_1 și X_2 sînt diagonal repartizate.

Propoziția 3.3.3. a) X_2 c.d.X_1 dacă și numai dacă (X_1, X_2) este diagonal repartizat în raport cu X_1.

b) X_1r.c.d.X_2 dacă și numai dacă X_1 și X_2 sînt diagonal repartizate.

Demonstrație. a) Dacă X_2 c.d.$X_1 \Rightarrow \sum_{x_2} p(x_2/x_1)\log p(x_2/x_1) = 0 \Leftrightarrow$ există un x_2^0 astfel că $p(x_2^0/x_1) = 1$ și $p(x_2/x_1) = 0$ pentru $x_2 \neq x_2^0$, ceea ce are loc dacă și numai dacă $p(x_1, x_2^0) = p(x_1)$ și $p(x_1, x_2) = 0$ pentru $x_2 \neq x_2^0$, deci dacă și numai dacă (X_1, X_2) este diagonal repartizat în raport cu X_1.

b) X_1r.c.d.$X_2 \Leftrightarrow X_1$c.d.X_2 și X_2c.d.X_1 și ca urmare a lui a) rezultă afirmația făcută.

Scanned with OKEN Scanner

Să observăm că

$$H(X_1) - H(X_1/X_2) = \sum_{x_1} \sum_{x_2} p(x_1, x_2) \log \frac{p(x_1, x_2)}{p(x_1)\, p(x_2)} \geqslant$$

$$\geqslant \sum_{x_1} \sum_{x_2} p(x_1, x_2) \left[1 - \frac{p(x_1)\, p(x_2)}{p(x_1, x_2)} \right] = 0$$

cu egalitate dacă și numai dacă $p(x_1, x_2) = p(x_1)p(x_2)$ pentru orice (x_1, x_2) deci dacă și numai dacă X_1 și X_2 sînt independente. Ca urmare a celor de mai sus putem scrie

Propoziția 3.3.4. Pentru orice vector aleator (X_1, X_2) avem

$$0 \leqslant H(X_1/X_2) \leqslant H(X_1) \tag{3.3.7}$$

cu primă egalitate dacă și numai dacă X_1c.d.X_2 și cu cea de a doua dacă și numai dacă X_1 și X_2 sînt independente.

Propoziția 3.3.5. Entropia Shannon $H(X_1, X_2)$ satisface următoarele :

(P. 1) este definită pentru orice vector aleator (X_1, X_2)

(P. 2) este simetrică în raport cu variabilele X_1 și X_2

(P. 3) $H(X_1, X_2) \geqslant 0$, cu egalitate dacă și numai dacă (X_1, X_2) este constant

(P. 4) $H(X_1, X_2) \geqslant H(X_1)$ cu egalitate dacă și numai dacă X_2c.d.X_1

(P.5) $H(X_1, X_2) \leqslant H(X_1) + H(X_2)$ cu egalitate dacă și numai dacă X_1 și X_2 sînt independente

(P.6) $H(X_1, X_2)$ este invariantă față de orice transformare biunivocă a variabilelor X_1 și X_2 .

Demonstrație. (P.1) — (P.4) și (P.6) sînt evidente, iar (P.5) rezultă din (3.3.5) și (3.3.7). Ca urmare deducem imediat

Propoziția 3.3.6. Pentru orice (X_1, X_2) avem

$$\min\{H(X_1), H(X_2)\} \leqslant \frac{1}{2} [H(X_1) + H(X_2)] \leqslant$$

$$\leqslant \max\{H(X_1), H(X_2)\} \leqslant H(X_1, X_2) \leqslant H(X_1) + H(X_2) \tag{3.3.8}$$

cu primele trei egalități simultan satisfăcute dacă și numai dacă X_1r.c.d.X_2 și cu ultima dacă și numai dacă X_1 și X_2 sînt independente

Ca o consecință a relației (3.3.5) putem scrie imediat

Scanned with OKEN Scanner

Propoziția 3.3.7. Pentru orice (X_1, X_2) avem

$$H(X_1) - H(X_1/X_2) = H(X_2) - H(X_2/X_1) \quad (3.3.9)$$

Relația (3.3.9) este cunoscută sub numele de „balanță informațională".

Pentru înțelegerea semnificației informaționale a relațiilor de mai sus, se impun unele precizări. Fiind date X_1 și X_2 vom nota cu

$H(X_1)$ = informația totală conținută în X_1 sau furnizată de X_1

$H(X_1/X_2)$ = informația proprie a lui X_1 după cunoașterea lui X_2.

Apar ca naturale afirmațiile

a) informația proprie a lui X_1 după cunoașterea lui X_2 coincide cu informația totală a lui X_1 dacă și numai dacă X_1 nu depinde de X_2, deci cînd ele sînt independente;

b) informația proprie a lui X_1 după cunoașterea lui X_2 este zero dacă și numai dacă X_1 este complet dependent de X_2

$H(X_1) - H(X_1/X_2)$ = informația pe care X_1 o furnizează ca urmare a cunoașterii lui X_2 sau informația pe care X_1 o „datorează" lui X_2.

De asemenea apar ca naturale afirmațiile

a) informația pe care X_1 o furnizează datorită cunoașterii lui X_2 coincide cu informația totală conținută de X_1 dacă și numai dacă X_1 c.d. X_2

b) informația pe care X_1 o furnizează datorită cunoașterii lui X_2 este zero dacă și numai dacă X_1 nu depinde de X_2, deci sînt independente.

Putem scrie:

$$\left\{\begin{array}{c}\text{Informația totală}\\ \text{furnizată de } X_1\\ H(X_1)\end{array}\right\} =$$

$$= \left\{\begin{array}{c}\text{Informația proprie}\\ \text{a lui } X_1 \text{ cînd se}\\ \text{cunoaște } X_2\\ H(X_1/X_2)\end{array}\right\} + \left\{\begin{array}{c}\text{Informația pe care}\\ X_1 \text{ o datorează cu-}\\ \text{noașterii lui } X_2\\ H(X_1) - H(X_1/X_2)\end{array}\right\} \quad (3.3.10)$$

Scanned with OKEN Scanner

Ca urmare, rapoartele $\frac{H(X_1/X_2)}{H(X_1)}$ și $\frac{H(X_1) - H(X_1/X_2)}{H(X_1)}$ reprezintă ponderea informației proprii a lui X_1 și respectiv datorată lui X_2 în informația totală a lui X_1. Cu aceste semnificații, balanța informațională (3.3.9) reprezintă faptul că „informația furnizată de X_1 prin cunoașterea lui X_2 este aceeași cu cea furnizată de X_2 prin cunoașterea lui X_1" sau că „ceea ce X_1 datorează lui X_2 este egal cu ceea ce X_2 datorează lui X_1", din punct de vedere al cantității de informație obținută după efectuarea experimentului sau al nedeterminării de dinainte de efectuarea experimentului.

Relația (3.3.5) poate fi citită sub forma „informația totală a cuplului (X_1, X_2) este egală cu informația totală a uneia din componente plus informația proprie a celeilalte".

Aceeași relație poate fi scrisă și sub forma

$$H(X_1, X_2) = H(X_1/X_2) + H(X_2/X_1) + \\ + [H(X_1) - H(X_1/X_2)] \qquad (3.3.11)$$

ceea ce înseamnă că „informația totală a cuplului este egală cu suma informațiilor proprii ale componentelor la care se adaugă informația datorată de una din ele celeilalte", sau sub forma

$$H(X_1, X_2) = H(X_1) + H(X_2) - [H(X_1) - H(X_1/X_2)] \qquad (3.3.12)$$

ceea ce înseamnă că „informația totală a cuplului este egală cu suma informațiilor totale ale componentelor din care scoatem informația pe care una din ele o datorează celeilalte".

Să remarcăm că, formal, relația (3.3.12) se aseamănă cu formula de însumare a probabilităților și că atît (3.3.11) cît și (3.3.12). apar ca naturale. Tot ca naturală apare și relația

$$H(X_1/X_2) + H(X_2/X_1) \leqslant H(X_1, X_2) \leqslant \\ \leqslant H(X_1 + H(X_2) \qquad (3.3.13)$$

a cărei demonstrație este imediată și care ar însemna că: „informația totală a cuplului este superioară sumei informațiilor proprii și inferioară sumei informațiilor totale ale componentelor cuplului".

Scanned with OKEN Scanner

Interpretările date pînă acum pot fi extinse fără dificultate și în cazul a mai mult de două variabile aleatoare, cu adaptarea textuală corespunzătoare.

3.3.2. Cazul tridimensional

Fie vectorul aleator tridimensional $X = (X_1, X_2, X_3)$, avînd repartiția $X : \begin{pmatrix} x \\ p(x) \end{pmatrix}$, $0 \leqslant p(x) \leqslant 1$, $\sum_x p(x) = 1$, $x = (x_1, x_2, x_3) \in A_1 \times A_2 \times A_3 \subset R_3$, unde A_1, A_2 și A_3 sînt mulțimile de valori, finite sau cel mult numărabile, ale variabilelor X_1, X_2, și X_3, iar $p(x) = p(x_1, x_2, x_3)$ este probabilitatea ca (X_1, X_2, X_3) să ia valoarea $(x_1, x_2\ x_3)$. Sînt cunoscute relațiile dintre repartiția lui X și repartițiile cuplurilor (X_j, X_k) sau variabilelor X_j, $1 \leqslant j \neq k \leqslant 3$. Avem deci $p(x_j, x_k) = \sum_{x_l} p(x_j, x_k, x_l)$ și $p(x_j) = \sum_{x_k} p(x_j, x_k) = \sum_{x_k} \sum_{x_l} p(x_j, x_k, x_l)$ cu $1 \leqslant j \neq k \neq l \leqslant 3$, fără a le mai scrie în continuare.

Definiția 3.3.9. Se numește entropie Shannon directă a vectorului (X_1, X_2, X_3) mărimea

$$H(X_1 X_2 X_3) = -\sum_{x_1} \sum_{x_2} \sum_{x_3} p(x_1, x_2, x_3) \log p(x_1, x_2, x_3) \tag{3.3.14}$$

Definiția 3.3.10. Se numește entropie a variabilei X_j condiționată de evenimentul $\{X_k = x_k, X_l = x_l\}$ expresia

$$H[X_j/(x_k, x_l)] = -\sum_{x_j} p(x_j/(x_k, x_l)) \log p(x_j/(x_k, x_l)) \tag{3.3.15}$$

Definiția 3.3.11. Se numește entropie a variabilei X_j condiționată de variabila X_k și de evenimentul $\{X_l = x_l\}$ expresia

$$H[X_j/X_k, x_l)] = \sum_{x_k} p(x_k)\, H[X_j/(x_k, x_l)] \tag{3.3.16}$$

și entropie a variabilei X_j condiționată de vectorul (X_k, X_l) expresia

$$H[X_j/(X_k, X_l)] = \sum_{x_k} \sum_{x_l} p(x_k, x_l)\, H[X_j/(x_k, x_l)] \tag{3.3.17}$$

Remarcăm că relația (3.3.17) mai poate fi pusă și sub forma

$$H[X_j/(X_k, X_l)] = -\sum_{x_j} \sum_{x_k} \sum_{x_l} p(x_j, x_k, x_l) \log p(x_j/(x_k, x_l)) \tag{3.3.18}$$

Scanned with OKEN Scanner

ţinînd cont că $p(x_j, x_k, x_l) = p(x_k, x_l)\, p(x_j/(x_k, x_l))$. Deducem

Propoziţia 3.3.8. Avem

$$H(X_1, X_2, X_3) = H(X_k, X_l) + H[X_j/(X_k, X_l)]$$

$$1 \leqslant j \neq k \neq l \leqslant 3 \qquad (3.3.19)$$

Definiţia 3.3.12. Se numeşte entropie a vectorului (X_j, X_k) condiţionat de evenimentul $\{X_l = x_l\}$, expresia

$$H[(X_j, X_k)/x_l] = -\sum_{x_j} \sum_{x_k} p((x_j, x_k)/x_l) \log p((x_j, x_k)/x_l) \qquad (3.3.20)$$

şi entropie a vectorului (X_j, X_k) condiţionat de variabila X_l, expresia

$$H[(X_j, X_k)/X_l] = \sum_{x_l} p(x_l) H[(X_j, X_k)/x_l] \qquad (3.3.21)$$

Relaţia (3.3.21) mai poate fi scrisă sub forma

$$H[(X_j, X_k)/X_l] = -\sum_{x_j} \sum_{x_k} \sum_{x_l} p(x_j, x_k, x_l) \log p((x_j, x_k)/x_l) \qquad (3.3.22)$$

deoarece $p(x_j, x_k, x_l) = p(x_l)\, p(x_j, x_k)/x_l)$. Deducem

Propoziţia 3.3.9. Avem

$$H(X_1, X_2, X_3) = H(X_l) + H[(X_j, X_k)/X_l],$$

$$1 \leqslant j \neq k \neq l \leqslant 3 \qquad (3.3.23)$$

Deoarece $p[(x_k, x_l)/x_j] = p(x_k/x_j).\ p[x_l/(x_j, x_k)]$, deducem că

$$H[(X_k, X_l)/X_j] = H(X_k/X_j) + H[X_l/(X_j, X_k)] \qquad (3.3.24)$$

şi ca urmare rezultă

Propoziţia 3.3.10. Avem

$$H(X_1, X_2, X_3) = H(X_j) + H(X_k/X_j) + H[X_l/(X_j, X_k)] \qquad (3.3.25)$$

Scanned with OKEN Scanner

Definiția 3.3.13. Spunem că vectorul (X_i, X_k) și variabila X_l sînt entropic independente dacă $H[(X_j, X_k)/X_l] = H(X_j, X_k)$ sau dacă $H[X_l/(X_j, X_k)] = H(X_l)$.

Din relațiile (3.3.24) și (3.3.25) se vede imediat că existența uneia din egalități implică existența celeilalte și ca urmare definiția poate fi dată doar cu una din ele. Ca și în cazul bidimensional se poate verifica ușor că independența entropică și cea probabilistică sînt echivalente.

Definiția 3.3.14. Spunem că

a) vectorul (X_j, X_k) este complet dependent de variabila X_l și scriem (X_j, X_k)c.d. X_l dacă $H[(X_j, X_k)/X_l] = 0$ și că X_lc.d.(X_j, X_k) dacă $H[X_l/(X_j, X_k)] = 0$

b) vectorul (X_j, X_k) și variabila X_l sînt reciproc complet dependente și scriem (X_j, X_k) r.c.d. X_l dacă

$$H[(X_j, X_k)/X_l] = H[X_l/(X_j, X_k)] = 0$$

c) vectorul (X_j, X_k) și variabila X_l sînt reciproc egal dependente dacă $H[(X_j, X_k)/X_l] = H[X_l(X_j, X_k)] \neq 0$.

Pornind de la definițiile 3.3.13 și 3.3.14 vom spune că X_1, X_2 și X_3 sînt independente în totalitate dacă orice două grupări ale acestora sînt independente și că ele sînt reciproc complet dependente în totalitate dacă orice două grupări ale lor sînt reciproc complet dependente, înțelegînd prin grupare o variabilă sau un cuplu de variabile din cele trei.

Propoziția 3.3.11. Avem

$$H[X_j/(X_k, X_l)] \leqslant H(X_j/X_k) \qquad (3.3.26)$$

cu egalitate dacă și numai dacă $p(x_j/(x_k, x_l)) = p(x_j/x_k)$ pentru orice (x_j, x_k, x_l).

Demonstrație. Inegalitatea rezultă din (3.3.24). Dar

$$H[X_j/(X_k, X_l)] - H(X_j/X_k) =$$

$$= \sum_{x_j} \sum_{x_k} \sum_{x_l} p(x_j, x_k, x_l) \log \frac{p(x_j/x_k)}{p(x_j/(x_k, x_l))} \leqslant$$

$$\leqslant \sum_{x_j} \sum_{x_k} \sum_{x_l} p(x_j, x_k, x_l) \left[\frac{p(x_j/x_k)}{p(x_j/(x_k, x_l))} - 1 \right] = 0 \quad \text{cu egalitate}$$

dacă și numai dacă $p(x_j/x_k) = p(x_j/(x_k, x_l))$ pentru toți (x_j, x_k, x_l).

Să remarcăm că dacă X_1, X_2 și X_3 sînt independente în totalitate, atunci are loc egalitatea (3.3.26) dar nu și reciproc. Egalitatea are de asemenea loc și în cazul cînd X_j este independentă de X_k și de (X_k, X_l).

Propoziția 3.3.12. Avem

$$H[(X_j, X_k)/X_l] \leqslant H(X_j, X_k) \tag{3.3.27}$$

cu egalitate dacă și numai dacă (X_j, X_k) și X_l sînt independente.

Demonstrație. $H(X_j, X_k) - H[(X_j, X_l)/X_k] =$

$$= \sum_{x_j} \sum_{x_k} \sum_{x_l} p(x_j, x_k, x_l) \log \frac{p((x_j, x_k)/x_l)}{p(x_j, x_k)} \geqslant$$

$$\geqslant \sum_{x_j} \sum_{x_k} \sum_{x_l} p(x_j, x_k, x_l) \left[1 - \frac{p(x_j, x_k)}{p((x_j, x_k)/x_l)}\right] = 0 \text{ cu egalitate}$$

dacă și numai dacă $p(x_j, x_k) = p((x_j, x_k)/x_l)$ ceea ce înseamnă că (X_j, X_k) și X_l sînt independente.

Propoziția 3.3.13. Avem

$$H(X_j/X_l) \leqslant H[(X_j, X_k)/X_l] \tag{3.3.28}$$

cu egalitate dacă și numai dacă X_k c.d. (X_j, X_l).

Demonstrație. Afirmația este consecință a relației (3.3.24). Din (3.3.26) — (3.3.28) rezultă

$$H[X_j/(X_k, X_l)] \leqslant H(X_j/X_l) \leqslant H[(X_j, X_k)/X_l] \leqslant$$

$$\leqslant H(X_j, X_k) \tag{3.3.29}$$

Propoziția 3.3.14. Dacă X_k c.d. X_l atunci

1. $H[X_j/(X_k, X_l)] = H(X_j/X_l)$
2. X_k c.d. (X_j, X_l)
3. $H[(X_j, X_k)/X_l] = H(X_j/X_l)$

Scanned with OKEN Scanner

4. $H[(X_k, X_l)/X_j] = H(X_l/X_j) = H(X_k/X_j) + H[X_l/(X_j, X_k)]$

5. $H[(X_j, X_l)/X_k] = H(X_j/X_l) + H(X_l/X_k)$

6. $H(X_j, X_k, X_l) = H(X_j, X_l)$.

Demonstrație. Din (3.3.25) rezultă că $H(X_k/X_l) + H[X_j/(X_k, X_l)] = H(X_j/X_l) + H[X_k/(X_j, X_l)]$. Deoarece X_k c.d. X_l rezultă că $H[X_j/(X_k, X_l)] = H(X_j/X_l) + H[X_k/(X_j, X_l)] \geqslant H(X_j/X_l)$ ceea ce contrazice (3.3.26). Deci $H[X_j/(X_k, X_l)] = H(X_j/X_l)$ și $H[X_k/(X_j, X_l)] = 0$ adică X_k c.d. (X_j, X_l). Punctele 3—5 rezultă ținînd cont și de (3.3.24) iar punctul 6 ținînd cont de (3.3.25).

Ca urmare a propoziției 3.3.14 deducem

Propoziția 3.3.15. Dacă X_k r.c.d. X_l atunci

1. $H[X_j/(X_k, X_l)] = H(X_j/X_k) = H(X_j/X_l)$

2. X_k c.d. (X_j, X_l) și X_l c.d. (X_j, X_k)

3. $H[(X_j, X_k)/X_l] = H(X_j/X_l)$

4. $H[(X_j, X_l)/X_k] = H(X_j/X_k)$

5. $H[(X_k, X_l)/X_j] = H(X_k/X_j) = H(X_l/X_j)$

6. $H(X_j, X_k, X_l) = H(X_j, X_k) = H(X_j, X_l)$

Propoziția 3.3.16. Dacă X_j c.d. X_k și X_k c.d. X_l atunci X_j c.d. X_l

Demonstrație. $H(X_j/X_l) = H[(X_j, X_k)/X_l] = H(X_k/X_l) + H[X_j/(X_k, X_l) = 0$ ca urmare a propoziției 3.3.15 și deci X_j c.d. X_l. Ca o consecință a propoziției 3.3.16 putem da

Propoziția 3.3.17. Totdeauna

1. Dacă X_j r.c.d. X_k și X_k r.c.d. X_l atunci X_j r.c.d X_l

2. X_j c.d. X_k c.d. X_l c.d. $X_j \Leftrightarrow$ oricare două sînt reciproc complet dependente.

Propoziția 3.3.18. Pentru orice vector aleator (X_1, X_2, X_3) avem

1. $H[X_j/(X_k, X_l)] = H(X_j/X_k) \Leftrightarrow H[X_k/(X_j, X_l)] = H(X_k/X_l)$

2. X_k c.d. $X_l \Leftrightarrow H[X_j/(X_k, X_l) = H(X_j/X_l)$ și $H(X_j, X_k, X_l) = H(X_j, X_l)$

Scanned with OKEN Scanner

3. X_k r.c.d. $X_l \Leftrightarrow \begin{cases} H[X_j/(X_k, X_l)] = H(X_j/X_k) = H(X_j/X_l) \\ H(X_j, X_k, X_l) = H(X_j, X_k) = H(X_j, X_l) \end{cases}$

4. (X_j, X_k) c.d. $X_l \Leftrightarrow X_j$ c.d. X_l și X_k c.d. X_l

Demonstrație. Afirmațiile 1 — 3 rezultă din (3.3.25) și propozițiile 3.3.14 și 3.3.15. Pentru cea de a patra, dacă (X_j, X_k) c.d. X_l, atunci $H(X_j, X_k, X_l) = H(X_l)$ și ca urmare $H(X_j/X_l) = H(X_k/X_l) = 0$, adică X_j c.d. X_l și X_k c.d. X_l. Invers, dacă X_j c.d. X_l atunci X_j c.d. (X_k, X_l) și cum X_k c.d. X_l avem $H(X_k/X_l) = 0$ și $H[X_j/(X_k, X_l)] = 0$ ceea ce implică $H(X_j, X_k, X_l) = H(X_l)$ deci (X_j, X_k) c.d. X_l.

Propoziția 3.3.19. Dacă X_j c.d. (X_k, X_l) și X_k c.d. X_l atunci X_j c.d. X_l.

Demonstrație. X_j c.d. (X_k, X_l) și X_k c.d. $X_l \Rightarrow H(X_j, X_k, X_l) = H(X_l)$ și X_k c.d. (X_j, X_l). Dar $H(X_j, X_k, X_l) = H(X_l) + H(X_j/X_l) + H[X_k/(X_j, X_l)] \Rightarrow H(X_j/X_l) = 0 \Rightarrow X_j$ c.d. X_l.

Propoziția 3.3.20. Entropia Shannon $H(X_1, X_2, X_3)$ satisface următoarele:

(P.1) este definită pentru orice vector aleator (X_1, X_2, X_3)

(P.2) este simetrică în raport cu variabilele X_1, X_2, X_3

(P.3) $H(X_1, X_2, X_3) \geqslant 0$, cu egalitate dacă și numai dacă (X_1, X_2, X_3) este constant

(P.4) $H(X_1, X_2, X_3) \geqslant H(X_j, X_l)$, cu egalitate dacă și numai dacă X_k c.d. (X_j, X_l), $1 \leqslant j \neq k \neq l \leqslant 3$

(P.5) $H(X_1, X_2, X_3) \geqslant H(X_j)$, cu egalitate dacă și numai dacă (X_k, X_l) c.d. X_j, $1 \leqslant j \neq k \neq l \leqslant 3$

(P.6) $H(X_1, X_2, X_3) \leqslant \sum_{j=1}^{3} H(X_j)$, cu egalitate dacă și numai dacă X_1, X_2 și X_3 sînt independente în totalitate

(P.7) $H(X_1, X_2, X_3)$ este invariantă pentru orice transformare biunivocă a variabilelor X_1, X_2 și X_3.

Demonstrație. (P.1) — (P.3) și (P.7) sînt evidente, iar (P.4) — (P.6) rezultă din (3.3.25). Ca urmare a propoziției 3.3.20 putem scrie

Propoziția 3.3.21. Avem totdeauna

$$1. \quad \min_{1 \leqslant j \leqslant 3} \{H(X_j)\} \leqslant \frac{1}{3} \sum_{j=1}^{3} H(X_j) \leqslant \max_{1 \leqslant j \leqslant 3} \{H(X_j)\} \leqslant$$

$$\leqslant H(X_1, X_2, X_3) \leqslant \sum_{j=1}^{3} H(X_j) \qquad (3.3.30)$$

Scanned with OKEN Scanner

cu primele trei egalităţi simultan dacă şi numai dacă $(X_1, X_2 X_3)$ sînt reciproc complet dependente în totalitate şi cu ultima dacă şi numai dacă ele sînt independente în totalitate

$$2. \quad \min_{1\leqslant j<k\leqslant 3}\{H(X_j, X_k)\} \leqslant \frac{1}{3}\sum_{j<k} H(X_j, X_k) \leqslant$$

$$\leqslant \max_{1\leqslant j<k\leqslant 3}\{H(X_j, X_k)\} \leqslant H(X_1, X_2, X_3) < \frac{2}{3}\sum_{j<k} H(X_j, X_k) \tag{3.3.31}$$

cu primele trei egalităţi dacă şi numai dacă fiecare variabilă este complet dependentă de celelalte două.

$$3. \quad \min_{j<k}\{H(X_j/X_k) + H(X_k/X_j) +$$

$$+ H[X_l/(X_j, X_k)]\} \leqslant \frac{1}{3}\sum_{j<k}[H(X_j/X_k) +$$

$$+ H(X_k/X_j) + H[X_l/(X_j, X_k)]] \leqslant \max_{j<k}\{H(X_j/X_k) +$$

$$+ H(X_k/X_j) + H[X_l/(X_j, X_k)]\} \leqslant$$

$$\leqslant H(X_1, X_2, X_3) \leqslant \frac{1}{3}\sum_{j,k=1}^{3} H(X_j/X_k) \tag{3.3.32}$$

cu egalitate peste tot dacă şi numai dacă (X_1, X_2, X_3) sînt independente în totalitate.

Demonstraţia rezultă din relaţia fundamentală (3.3.25). Aplicaţii deosebite în capitolele următoare are relaţia (3.3.31).

Tot din relaţia fundamentală (3.3.25) putem scrie relaţii de balanţă informaţională de tipul (3.3.9).

Propoziţia 3.3.22. Avem totdeauna

$$1. \quad H(X_j) - H[X_j/(X_k, X_l)] =$$

$$= H(X_k) - H(X_k/X_j) + H(X_l/X_k) -$$

$$- H[X_l/(X_j, X_k)] \tag{3.3.33}$$

$$2. \quad H(X_j/X_l) - H[(X_j/(X_k, X_l)] =$$

$$= H(X_k/X_l) - H[X_k/(X_j, X_l)] \tag{3.3.34}$$

Scanned with OKEN Scanner

Dacă notăm $H[(X_j, X_k)/X_l] = H(X_j/X_l, X_k/X_l)$, atunci relația (3.3.34) apare ca o balanță informațională de tipul (3.3.9) pentru cuplul aleator $(X_j/X_l, X_k/X_l)$.

3.3.3. Cazul n-dimensional

Fie vectorul aleator discret n-dimensional $X = (X_1, X_2, \ldots, X_n)$ adică $X : \binom{x}{p(x)}$, $x = (x_1, x_2, \ldots, x_n) \in \prod_{k=1}^{n} A_k \subset \mathbf{R}^n$, A_k fiind mulțimea de valori ale variabilei aleatoare X_k, $1 \leqslant k \leqslant n$, iar $p(x) = p(x_1, \ldots, x_n)$ probabilitatea ca $(X_1, X_2, \ldots, X_n)$ să ia valoarea $(x_1, x_2, \ldots, x_n)$. Sînt cunoscute modalitățile de însumare, pornind de la repartiția comună $p(x_1, x_2, \ldots, x_n)$, pentru a obține repartițiile comune sau individuale a oricare m variabile din cele n, $1 \leqslant m \leqslant n$, ca extensii naturale ale celor scrise pentru $n = 3$.

Definiția 3.3.15. Se numește entropie Shannon directă a vectorului aleator $X = (X_1, X_2, \ldots, X_n)$ expresia

$$H(X) = H(X_1, X_2, \ldots, X_n) =$$

$$= -\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} p(x_1, \ldots, x_n)\log\, p(x_1, \ldots, x_n) \qquad (3.3.35)$$

În mod asemănător definițiilor 3.3.10 — 3.3.12 putem da expresiile entropiilor condiționate și în cazul n-dimensional. Dacă notăm cu Y_1, Y_2 și Y_3 trei vectori aleatori disjuncți formați cu componente ale lui X și cu y_1, y_2 și y_3 vectorii valori luate de aceștia, atunci prin analogie cu definițiile 3.3.10 — 3.3.12 putem scrie expresiile pentru $H[Y_1/(y_2, y_3)]$, $H[Y_1/(Y_2, y_3)]$, $H[Y_1/(Y_2, Y_3)]$, $H[(Y_1, Y_2)/y_3]$ sau $H[(Y_1, Y_2)/Y_3]$, formal aceleași ca în definițiile amintite, dar cu altă semnificație ca urmare a faptului că Y_1, Y_2 și Y_3 sînt vectori aleatori. După cum Y_1, Y_2 sau Y_3 sînt variabile aleatoare obținem diverse cazuri particulare de entropii condiționate, între care și ceea ce s-a prezentat la cazul tridimensional.

Definiția 3.3.16. Spunem că

1. vectorul aleator Y_1 este complet dependent de vectorul Y_2 dacă $H(Y_1/Y_2) = 0$

2. vectorii Y_1 și Y_2 sînt reciproc complet dependenți dacă $H(Y_1/Y_2) = H(Y_2/Y_1) = 0$

Scanned with OKEN Scanner

3. vectorii Y_1 și Y_2 sînt egal dependenți unul de altul dacă $H(Y_1/Y_2) = H(Y_2/Y_1) \neq 0$.

Observăm că definițiile 3.3.5 și 3.3.16 sînt formal identice, ultima generalizînd chiar definiția 3.3.14.

Vom spune că $X_1, X_2, \ldots, X_n$ sînt reciproc complet dependente în totalitate dacă orice doi vectori formați cu componente diferite ale lui X sînt reciproc complet dependenți.

Definiția 3.3.17. Spunem că Y_1 și Y_2 sînt independenți dacă $H(Y_1/Y_2) = H(Y_1)$.

Vom spune că $X_1, X_2, \ldots, X_n$ sînt independente în totalitate dacă orice doi vectori Y_1 și Y_2 sînt independenți. Am precizat în paragrafele precedente că independența entropică și independența probabilistică sînt echivalente, afirmație valabilă și în continuare, cu demonstrație asemănătoare.

Din relația (3.3.24) scrisă pentru vectorii Y_1, Y_2 și Y_3 deducem că

$$H[(X_{i_2}, \ldots, X_{i_1})/X_{i_1}] = \sum_{l=2}^{k} H[X_{i_l}/(X_{i_1}, \ldots, X_{i_{l-1}})] \tag{3.3.36}$$

și ca urmare avem,

$$H(X_1, X_2, \ldots, X_n) = H(X_{i_1}) + \sum_{k=2}^{n} H[X_{i_k}/(X_{i_1}, \ldots, X_{i_{k-1}})] \tag{3.3.37}$$

relație care generalizează (3.3.25).

Generalizînd (3.3.26) putem scrie că

$$H[X_{i_k}/(X_{i_1}, \ldots, X_{i_{k-1}})] \leqslant H(X_{i_k}/X_{i_{k-1}}) \leqslant H(X_{i_k}) \tag{3.3.38}$$

cu primă egalitate dacă $X_{i_1}, X_{i_2}, \ldots, X_{i_{k-2}}$ sînt complet dependente de X_{k-1} și cu ultima egalitate dacă și numai dacă $X_{i_1}, X_{i_2}, \ldots, X_{i_{k-1}}$ sînt independente în totalitate. Din (3.3.37) și (3.3.38) rezultă

$$H(X_1, \ldots, X_n) \leqslant H(X_{i_1}) +$$

$$+ \sum_{k=2}^{n} H[X_{i_k}/(X_{i_1}, \ldots, X_{i_{k-1}})] \leqslant H(X_{i_1}) + \sum_{k=2}^{n} H(X_{i_k}/X_{i_{k-1}}) \tag{3.3.39}$$

pentru orice permutare $\{i_1, i_2, \ldots, i_n\}$ a mulțimii $\{1, 2, \ldots, n\}$.

Propozițiile 3.3.11—3.3.19 pot fi formulate și deduse cu ușurință și în cazul vectorilor aleatori Y_1, Y_2 și Y_3,

Scanned with OKEN Scanner

căpătînd semnificații deosebite cînd unul sau doi dintre vectori se reduc la variabile aleatoare. Astfel, ținînd cont și de (3.3.39) deducem că

$$H[(X_1, X_2, \ldots, X_k)/(X_{k+1}, \ldots, X_m)] =$$

$$= \sum_{j=1}^{k} H[X_j/(X_1, \ldots, X_{j-1}; X_m)] - \qquad (3.3.39')$$

$$- \sum_{j=k+1}^{m} H[X_j/(X_{k+1}, \ldots, X_{j-1}; X_m)]$$

Propoziția 3.3.23. Variabilele $X_1, X_2, \ldots, X_n$ sînt reciproc complet dependente în totalitate dacă și numai dacă pentru orice $1 \leqslant j \neq k \leqslant n$, X_j r.c.d. X_k.

Demonstrație. Dacă $(X_1, X_2, \ldots, X_n)$ sînt reciproc complet dependente în totalitate, atunci orice doi vectori formați cu variabile diferite sînt reciproc complet dependenți și în particular orice două variabile. Invers, fie vectorii $Y_1 = (X_1, \ldots, X_k)$ și $Y_2 = (X_{k+1}, \ldots, X_m)$, $1 \leqslant k < m \leqslant n$. Avem

$$H(Y_1/Y_2) = \sum_{j=2}^{k} H[X_j/(X_1, \ldots, X_{j-1}; Y_2) = 0$$

în baza propoziției 3.3.14 și analog pentru $H(Y_2/Y_1)$. Deci Y_1 r.c.d. Y_2 și ca urmare $(X_1, X_2, \ldots, X_n)$ sînt reciproc complet dependente în totalitate.

Propoziția 3.3.24. Dacă $X_1, X_2, \ldots, X_n$ sînt independente în totalitate atunci oricare dintre ele este independentă de orice vector format din celelalte variabile și reciproc.

Demonstrație. Dacă $(X_1, \ldots, X_n)$ sînt independente în totalitate, atunci $H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i)$ și ținînd cont de (3.3.37) rezultă că pentru orice permutare $(i_1, i_2, \ldots, i_n)$ avem $H[X_{i_k}|(X_{i_1}, \ldots, X_{i_{k-1}})] = H(X_{i_k})$ și ca urmare prima afirmație. Reciproc, dacă fiecare X_{i_k} este independentă de orice grupare a celorlalte, atunci în particular este independentă și de $(X_{i_1}, \ldots, X_{i_{k-1}})$ și ca urmare $H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i)$, deci $X_1, X_2, \ldots, X_n$ sînt independente în totalitate.

Mai general, putem deduce

Propoziția 3.3.25. Dacă $X = (X_1, \ldots, X_n)$ sînt independente în totalitate atunci oricare ar fi $Y_1 \subset X$, $Y_2 \subset X$, $Y_1 \cap Y_2 = \emptyset$, ei sînt independenți și reciproc.

Scanned with OKEN Scanner

Ținînd cont de cele de mai sus, putem formula o generalizare a propoziției 3.3.20 și ca urmare putem scrie

Propoziția 3.3.26. Avem

$$\min_{1\leqslant j\leqslant n}\{H(X_j)\} \leqslant \frac{1}{n}\sum_{j=1}^{n} H(X_j) \leqslant$$

$$\leqslant \max_{1\leqslant j\leqslant n}\{H(X_j)\} \leqslant H(X_1,\ldots,X_n) \leqslant \sum_{j=1}^{n} H(X_j) \quad (3.3.40)$$

cu egalitate peste tot în primele trei inegalități dacă $(X_1,\ldots X_n,)$ sînt reciproc complet dependente în totalitate și cu ultima egalitate dacă ele sînt independente în totalitate.

Propoziția 3.3.27. Există totdeauna numerele $\alpha_{i_k}\in[0,1]$, $1\leqslant\sum_{k=1}^{n}\alpha_{i_k}\leqslant n$, astfel că

$$H(X_1,\ldots,X_n)=\sum_{k=1}^{n}\alpha_{i_k}H(X_{i_k}) \quad (3.3.41)$$

pentru orice permutare a indicilor $\{1,2,\ldots,n\}$.

Demonstrație. Este suficient să luăm $\alpha_{i_k}=\dfrac{H[X_{i_k}/(X_{i_1},\ldots,X_{i_{k-1}})]}{H(X_{i_k})}$, caz în care α_{i_k} reprezintă ponderea informației proprii a lui X_{i_k} cînd se cunosc X_{i_1}, $X_{i_2},\ldots,X_{i_{k-1}}$ în informația totală a lui X_{i_k}.

Fie acum $Y=(Y_j)_{j=1}^{m}$, $Y_j\cap Y_k=\emptyset$, $j\neq k$, o partiție a lui X, ceea ce ar însemna considerarea a m grupuri de experimente din cele n aflate în discuție (experimente sau factori economico-sociali ai unui anumit proces sau fenomen economico-social, sau caracteristici ale unei populații statistice etc.). Există suficiente exemple practice care să ilustreze considerarea unor astfel de situații, încît fiecare Y_j să conțină cel puțin o variabilă X_i. Notînd cu y_j vectorul valoare al lui Y_j și cu $y=(y_1, y_2,\ldots,y_m)$ vectorul valoare al lui Y și presupunînd cunoscută repartiția comună $p(y)=p(y_1, y_2,\ldots,y_m)$ a vectorilor $Y_1, Y_2,\ldots,Y_m$, putem deduce repartițiile individuale $p(y_i)$ ca fiind $p(y_j)=\sum_{\substack{y_k\\ 1\leqslant k\neq j\leqslant m}} p[(y_k)_{k=1}^{m}]$, unde $p[(y_k)_{k=1}^{m}]=p(y_1,\ldots,y_m)$. Pentru fiecare Y_j entropia poate fi determinată cu ajutorul relației (3.3.35) și orice entropie condiționată poate fi scrisă în virtutea relațiilor 3.3.15—3.3.24, adaptate noilor notații.

Scanned with OKEN Scanner

Definiția 3.3.18. Se numește entropie Shannon directă a vectorului Y expresia

$$H(Y) = -\sum_{y} p(y) \log p(y) =$$

$$= -\sum_{y_1} \cdots \sum_{y_m} p(y_1, \ldots, y_m (\log p(y_1, \ldots, y_m) =$$

$$= H(Y_1, \ldots, Y_m) \tag{3.3.42}$$

Să remarcăm că relația (3.3.42) generalizează relațiile (3.3.1), (3.3.14) și (3.3.35) care pot fi obținute din aceasta pentru $m = n = 2$, $m = n = 3$ și $m = n$.

Propoziția 3.3.28. Pentru orice partiționare $Y = (Y_j)_{j=1}^{m}$ a lui $X = (X_i)_{i=1}^{n}$ are loc relația

$$H(Y) = H(X) \tag{3.3.43}$$

Demonstrație. Pentru ușurința scrierii să presupunem că $m = 2$ și că $Y_1 = (X_i)_{i=1}^{k}$ și $Y_2 = (X_i)_{i=k+1}^{n}$. Avem $H(X) = H(Y_1, Y_2) = H(Y_1) + H(Y_2/Y_1)$. Dar $H(Y_1) = H(X_1) + \sum_{i=2}^{k} H[X_i/(X_1, \ldots, X_{i-1})]$ în baza relației (3.3.37) iar $H(Y_2/Y_1) = \sum_{i=k+1}^{m} H[X_i/(X_1, \ldots, X_k; X_{k+1}, \ldots, X_{i-1})]$ în baza relației (3.3.36) și ca urmare

$$H(Y) = H(X_1) + \sum_{i=2}^{m} H[X_i/(X_1, \ldots, X_{i-1})] = H(X)$$

în baza relației (3.3.37). Într-o scriere mai complicată putem da (3.3.43) pentru orice m.

Ca urmare a relației (3.3.43), aparent simplă, rezultă că indiferent cum s-ar aranja elementele $X_1, \ldots, X_n$ în grupurile $Y_1, \ldots, Y_m$ informația totală a ansamblului format cu $X_1, X_2, \ldots, X_n$ sau cu grupurile $Y_1, Y_2, \ldots, Y_m$ este aceeași. Cu alte cuvinte putem scrie că măsura cantității de informație sau a nedeterminării unui grup de n experimente este invariantă în raport cu orice grupare a experimentelor. Următoarea propoziție generalizează toate rezultatele anterioare cu privire la informația totală a unui grup de experimente sau a unui vector aleator și rezultă ca o sinteză a celor de mai sus.

Scanned with OKEN Scanner

Propoziția 3.3.29. Entropia Shannon $H(Y_1, Y_2, \ldots, Y_m)$ satisface următoarele

(P.1) este definită pentru orice partiție $Y = (Y_j)_{j=1}^m$ a lui X

(P.2) este invariantă în raport cu partiția Y a lui X, adică

$$H(Y) = H(X) \text{ oricare ar fi partiția } Y = (Y_j)_{j=1}^m$$

(P.3) este simetrică în raport cu elementele partiției $Y = (Y_j)_{j=1}^m$

(P.4) $H(Y) \geqslant H(Z)$ oricare ar fi $Z \subseteq Y$ cu egalitate dacă și numai dacă $(Y \setminus Z)$ c.d. Z

(P.5) $H(Y_1, \ldots, Y_m) = H(Y \setminus Y_j)$ dacă și numai dacă există un $j \in \{1, 2, \ldots, m\}$ astfel că Y_j c.d. $(Y \setminus Y_j)$

(P.6) $H(Y_1, \ldots, Y_m) = H(Y_j)$ dacă și numai dacă există un $j \in \{1, 2, \ldots, m\}$ astfel că $(Y \setminus Y_j)$ c.d. Y_j

(P.7) $H(Y_1, \ldots, Y_m) = H(Z) = H(Y \setminus Z)$ dacă și numai dacă există $Z \subset Y$ astfel încît Z r.c.d. $(Y \setminus Z)$

(P.8) $H(Y_1, \ldots, Y_m) \leqslant H(Z) + H(Y \setminus Z)$ pentru orice $Z \subset Y$, cu egalitate dacă și numai dacă Z și $(Y \setminus Z)$ sînt independenți

(P.9) $H(Y_1, \ldots, Y_m) \leqslant \sum_{j=1}^{m} H(Y_j) \leqslant \sum_{i=1}^{n} H(X_i)$ cu prima egalitate dacă și numai dacă $(Y_j)_{j=1}^m$ sînt independenți în totalitate și cu egalitate peste tot dacă și numai dacă $(X_i)_{i=1}^n$ sînt independente în totalitate

(P.10) este invariantă în raport cu orice transformare biunivocă a elementelor partiției $(Y_j)_{j=1}^m$.

Ca o consecință imediată a propoziției (3.3.29) putem scrie

Propoziția 3.3.30. Avem intotdeauna

$$\min_{1 \leqslant j \leqslant m} \{H(Y_j)\} \leqslant \frac{1}{m} \sum_{j=1}^{m} H(Y_j) \leqslant \max_{1 \leqslant j \leqslant m} \{H(Y_j)\} \leqslant H(Y) \leqslant \sum_{j=1}^{m} H(Y_j) \tag{3.3.44}$$

cu primele trei egalități simultan satisfăcute dacă și numai dacă $(Y_j)_{j=1}^m$ sînt reciproc complet dependente în totalitate și cu ultima dacă și numai dacă $(Y_j)_{j=1}^m$ sînt independenți în totalitate.

Scanned with OKEN Scanner

Ținînd cont de propoziția 3.3.27 putem formula

Propoziția 3.3.31. Există totdeauna numerele $\alpha_{jk} \in [0, 1]$ astfel că

$$H(Y_1, \ldots, Y_m) = \sum_{j=1}^{m} \sum_{k=1}^{s_j} \alpha_{jk} H(X_k^j) \qquad (3.3.45)$$

unde $Y_j = (X_k^j)_{k=1}^{s_j}$, $\sum_{j=1}^{m} s_j = n$.

Demonstrație. Într-adevăr, ținînd cont de relația (3.3.37) scrisă pentru Y și apoi pentru fiecare Y_j, este suficient să luăm

$$\alpha_{jk} = \frac{H[Y_j/(Y_1, \ldots, Y_{j-1})]}{H(Y_j)} \cdot \frac{H[X_k^j/(X_1^j, \ldots, X_{k-1}^j)]}{H(X_k^j)}$$

$$1 \leqslant k \leqslant s_j, \quad 1 \leqslant j \leqslant m \qquad (3.3.46)$$

și afirmația este verificată.

Dacă ținem cont și de relația (3.3.43), rezultă că relația (3.3.45) dă o altă posibilitate de însumare ponderată a informațiilor totale ale componentelor lui X decît cea dată de relația (3.3.41), urmare a invarianței la partiționare a lui X.

3.4. Entropii ale unui vector aleator continuu

Fie $X = (X_j)_{i=1}^{n}$ un vector aleator continuu adică, $X : \binom{x}{p(x)}$, $x \in D \subseteq \mathbf{R}^n$, $p(x) = p(x_1, \ldots, x_n) \geqslant 0$, $\int_D p(x)\mathrm{d}x = 1$, $\mathrm{d}x = \prod_{i=1}^{n} \mathrm{d}x_i$. Ca o generalizare a rezultatelor din paragrafele 3.2 și 3.3, vom face cîteva considerații asupra entropiei în cazul unui vector aleator continuu, fără însă a insista prea mult asupra acestor chestiuni, deoarece în practică, mai ales în măsurarea dependenței dintre două sau mai multe variabile aleatoare, repartițiile continue sînt mai puțin frecvente.

Pornind de la repartiția comună $p(x_1, x_2, \ldots, x_n)$ a variabilelor $X_1, X_2, \ldots, X_n$ putem obține repartițiile individuale $p(x_i)$ ale variabilelor X_i, $1 \leqslant i \leqslant n$, sau repartițiile unor vectori aleatori formați cu două sau mai multe dintre cele n variabile. Avem spre exemplu $p(x_i) = \int_D p(x) \prod_{\substack{k=1 \\ k \neq 1}}^{n} \mathrm{d}x_k$ sau

Scanned with OKEN Scanner

$p(x_i, x_j) = \int_D p(x) \prod_{\substack{k=1 \\ k \neq i,j}}^{n} dx_k$, pentru orice $1 \leqslant i \neq j \leqslant n$ şi în mod analog pentru mai mult de două variabile.

Definiţia 3.4.1. Se numeşte entropie Shannon directă a vectorului aleator continuu X expresia

$$H(X) = -\int_D p(x) \log p(x)\, dx = -\int_D \dots \int p(x_1, x_2, \dots$$

$$\dots, x_n) \log p(x_1, x_2, \dots, x_n) \cdot dx_1 \dots dx_n = H(X_1, X_2, \dots, X_n) \tag{3.4.1}$$

Remarcăm că relaţia (3.4.1) generalizează relaţiile (3.2.1) şi (3.3.35), fără să fie însă o trecere la limită în cazul ultimei dintre acestea. Să considerăm acum vectorii aleatori Y_1, Y_2 şi Y_3 formaţi cu componente ale lui X. Analog definiţiilor (3.3.10) — (3.3.12) putem scrie şi în cazul continuu expresiile diverselor entropii condiţionate.

Definiţia 3.4.2. Se numeşte entropie a vectorului Y_1 condiţionat de evenimentul $\{(Y_2, Y_3) = (y_2, y_3)\}$ expresia

$$H[Y_1/(y_2, y_3)] = -\int_{D(y_1)} p[y_1/(y_2, y_3)] \log [p[y_1/(y_2, y_3)]\, dy_1 \tag{3.4.2}$$

entropie a vectorului Y_1 condiţionat de vectorul Y_2 şi evenimentul $\{Y_3 = y_3\}$ expresia

$$H[Y_1/(Y_2, y_3)] = \int_{D(y_2)} p(y_2)\, H[Y_1/(y_2, y_3)]\, dy_2 \tag{3.4.3}$$

şi entropie a vectorului Y_1 condiţionat de cuplul (Y_2, Y_3) expresia

$$H[Y_1/(Y_2, Y_3)] = \int_{D(y_2, y_3)} p(y_2, y_3)\, H[Y_1/(y_2, y_3)]\, dy_2 dy_3 \tag{3.4.4}$$

unde $D(y_1)$, $D(y_2)$ şi $D(y_2, y_3)$ sînt domeniile de valori ale vectorilor Y_1, Y_2 şi (Y_2, Y_3).

Definiţia 3.4.3. Se numeşte entropie a cuplului (Y_1, Y_2) condiţionat de evenimentul $\{Y_3 = y_3\}$, expresia

$$H[(Y_1, Y_2)/y_3] = -\int_{D(y_1,y_2)} p[(y_1, y_2)/y_3] \log p[(y_1, y_2)/y_3]\, dy_1\, dy_2 \tag{3.4.5}$$

şi condiţionat de vectorul Y_3, expresia

$$H[(Y_1, Y_2)/Y_3] = \int_{D(y_3)} p(y_3) H[(Y_1, Y_2)/y_3]\, dy_3 \tag{3.4.6}$$

Cu aceste precizări putem formula definiţiile 3.3.16 şi 3.3.17 precum şi unele dintre relaţiile din paragraful 3.3, cum ar fi (3.3.36) — (3.3.45), care rămîn valabile şi în cazul continuu, motiv pentru care nu le mai transcriem. Vom remarca de asemenea că propoziţia 3.2.1 rămîne valabilă şi în cazul vectorial.

Propoziţia 3.4.1. Dacă $X = (X_1, X_2)$ are densitatea de repartiţie $p(x_1, x_2) : \mathbf{R}^2 \to [0, \infty)$

$$p(x_1, x_2) = \frac{\sqrt{a_1 a_2 - h^2}}{2\pi} \exp\left\{-\frac{1}{2} a_1 (x_1 - \lambda_1)^2 + \right.$$
$$\left. + 2h(x_1 - \lambda_1)(x_2 - \lambda_2) + a_2 (x_2 - \lambda_2)^2\right\} \tag{3.4.7}$$

unde $(\lambda_1, \lambda_2) \in \mathbf{R}^2$, $(a_1, a_2) \in \mathbf{R}_+^2$ şi $h \in \mathbf{R}$ sînt constante cu $a_1 a_2 - h^2 > 0$, atunci

1) $H(X_i) = \log \sqrt{\dfrac{2\pi e\, a_j}{a_1 a_2 - h^2}}$;

2) $H(X_i/X_j) = \log \sqrt{\dfrac{2\pi e}{a_i}} \quad 1 \leqslant i \neq j \leqslant 2$

3) $H(X_1, X_2) = \log \dfrac{2\pi e}{\sqrt{a_1 a_2 - h^2}}$

Demonstraţie. Deducem mai întîi că

$$p(x_i) = \sqrt{\frac{a_1 a_2 - h^2}{2\pi a_j}} \exp\left\{-\frac{a_1 a_2 - h^2}{2 a_j}(x_i - \lambda_i)^2\right\}$$

$$p(x_i/x_j) = \sqrt{\frac{a_i}{2\pi}} \exp\left\{-\frac{1}{2}\left[\sqrt{a_i}(x_i - \lambda_i) + \frac{h(x_j - \lambda_j)}{\sqrt{a_i}}\right]^2\right\}$$

Scanned with OKEN Scanner

și ca urmare, prin calcul direct obținem 1) și 2). $H(X_1, X_2)$ poate fi obținută direct sau ținînd cont de 1) și 2) ca urmare a relației (3.3.5) valabilă și în cazul continuu. La fel obținem pe oricare din ele cînd știm pe celelalte două. Aceleași rezultate le obținem și în cazul particular $\lambda_1 = \lambda_2 = 0$, parametrii λ_1 și λ_2 fiind valorile medii ale variabilelor X_1 și X_2, fapt ce poate fi dedus imediat.

Dacă luăm cazul particular al repartiției (3.4.7) în care :

$$\lambda_1 = \lambda_2 = 0,\ a_i = \frac{1}{\sigma_i^2(1-\rho^2)} \text{ și } h = \frac{-\rho}{\sigma_1\sigma_2(1-\rho^2)},$$

$i = 1, 2$ unde σ_i este abaterea medie pătratică a variabilei X_i iar ρ coeficientul de corelație clasic al variabilelor X_1 și X_2 atunci

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times$$

$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x_1^2}{\sigma_1^2} - \frac{2\rho x_1 x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2}\right)\right\} \quad (3.4.8)$$

reprezintă densitatea de repartiție normală bidimensională $N(0, 0, \sigma_1, \sigma_2, \rho)$. Ca urmare deducem imediat

Propoziția 3.4.2.. Dacă $X = (X_1, X_2)$ este repartizat normal bidimensional, atunci

1) $H(X_i) = \log(\sqrt{2\pi e}\,\sigma_i)$;

2) $H(X_i/X_j) = \log[\sqrt{2\pi e(1-\rho^2)}\,\sigma_i]$; $1 \leqslant i \neq j \leqslant 2$

3) $H(X_1, X_2) = \log(2\pi e\,\sigma_1\sigma_2\sqrt{1-\rho^2})$

Rezultatele de mai sus pot fi extinse la repartiția n-dimensională a cărei densitate este

$$p(x) = \frac{1}{(2\pi)^{n/2}}\sqrt{|A|}\exp\left\{-\frac{1}{2}(x-\lambda)A(x-\lambda)'\right\} \quad (3.4.9)$$

unde A este o matrice pătratică simetrică, cu elementele de pe diagonala principală precum și cu minorii acestora de semn pozitiv, iar λ un vector constant real. Se poate arăta că repartiția marginală a variabilei X_i este

$$p(x_i) = \sqrt{\frac{|A|}{2\pi M_{ii}}}\exp\left\{-\frac{1}{2}\frac{|A|}{M_{ii}}(x_i - \lambda_i)^2\right\} \quad 1 \leqslant i \leqslant n \quad (3.4.10)$$

Scanned with OKEN Scanner

iar dacă Y_k este un vector al partiției $Y=(Y_k)_{k=1}^m$ a lui X, format din n_k componente ale lui X, $\sum_{k=1}^{m} n_k = n$, atunci

$$p(y_k) = \frac{\sqrt{|A_k|}}{(2\pi)^{n_k/2}} \exp\left\{-\frac{1}{2}(y_k - \mu_k) A_k (y_k - \mu_k)'\right\} \tag{3.4.11}$$

unde A_k este o matrice pătratică de dimensiune n_k cu aceleași proprietăți ca și A din (3.4.9), $|A_k|$ fiind determinantul lui A_k, ca și $|A|$ pentru A, iar μ_k o parte a lui λ corespunzătoare celor n_k componente ale lui X ce formează pe Y_k. În aceste condiții găsim

1) $H(X_i) = \frac{1}{2} \log \frac{2\pi e M_{ii}}{|A|}$;

2) $H(Y_k) = \frac{1}{2} \log \frac{(2\pi e)^{n_k}}{|A_k|}$;

3) $H(X) = \frac{1}{2} \log \frac{(2\pi e)^n}{|A|}$.

3.5. Unicitatea entropiei Shannon. Extensiuni

Definind entropia ce îi poartă numele, Shannon a propus un sistem de axiome considerate ca fiind foarte naturale și deductibile în mod intuitiv, pe care trebuie să le satisfacă o măsură a cantității de informație. Axiomatica lui Shannon a fost continuată de altele, scopul fiind același, de a dovedi că măsura cantității de informație $H(X)$, definită de Shannon în 1948, este unic determinată, atît pentru variabile aleatoare discrete cît și pentru cele continue, axiomatici pe care le vom prezenta în continuare fără demonstrațiile corespunzătoare, acestea fiind riguros expuse în diferite lucrări ce abordează această temă.

3.5.1. Axiomatica lui C. E. Shannon (1948)

Propoziția 3.5.1. Fie $\{H_n(p_1, p_2, \ldots, p_n)\}_{n \in \mathbf{N}}$ un șir de funcții nenegative definite pe domeniul $p_k \geqslant 0$, $1 \leqslant k \leqslant n$ $\sum_{K=1}^{n} p_k = 1$, $n \in \mathbf{N}$.

Scanned with OKEN Scanner

Dacă

(S.1) $H_2(1/2,\ 1/2) = 1$

(S.2) $H_n(p_1,\ p_2,\ldots,p_n)$ este o funcţie continuă şi simetrică în raport cu toate argumentele, oricare ar fi $n \in \mathbf{N}$

$$(S.3)\quad H_n(p_1,\ldots,p_{n-1},\ p_m q_{m,m},\ p_m q_{m,m+1},\ldots,p_m q_{m,n}) =$$
$$= H_m(p_1,\ p_2,\ldots,p_m) + p_m H_{n-m+1}(q_{m,m},\ q_{m,m+1},\ldots,q_{m,n})$$

unde $n > m > 1$, $p_1 + p_2 + \ldots + p_m = 1$; $q_{m,m} + q_{m,m+1} + \ldots + q_{m,n} = 1$

$$(S.4)\quad H_n\left(\frac{1}{n},\ldots,\frac{1}{n}\right) \leqslant H_{n+1}\left(\frac{1}{n+1},\ldots,\frac{1}{n+1}\right)$$

atunci $H_n(p_1,\ldots,p_n)$ este determinată în mod unic de relaţia

$$H_n(p_1,\ p_2,\ldots,\ p_n) = -\sum_{k=1}^{n} p_k \log p_k \qquad (3.5.1)$$

3.5.2. Axiomatica lui A. I. Hincin (1956)

Propoziţia 3.5.2. Fie $\{H_n(p_1,\ p_2,\ \ldots,p_n)\}_{n\in\mathbf{N}}$ un şir de funcţii nenegative definite pe domeniul $p_k \geqslant 0$, $1 \leqslant k \leqslant n$, $\sum_{k=1}^{n} p_k = 1$, $n \in \mathbf{N}$.

Dacă

(H.1) $H_n(p_1,\ldots,p_n)$ este o funcţie continuă şi simetrică în raport cu toate argumentele oricare ar fi $n \in N$

(H.2) $H_n(p_1,\ldots,\ p_n,\ 0) = H_n(p_1,\ldots,p_n)$ pentru orice $n \in \mathbf{N}$

(H.3) $H_n(p_1,\ldots,p_n) \leqslant H_n\left(\frac{1}{n},\ldots,\frac{1}{n}\right)$ pentru orice $n \in \mathbf{N}$

$$(H.4)\quad H_{nm}(p_{11},\ldots,p_{nm}) = H_n(p_1,\ldots,p_n) +$$
$$+ \sum_{k=1}^{n} p_k H_m\left(\frac{p_{k1}}{p_k},\ldots,\frac{p_{km}}{p_k}\right)$$

unde $p_{kl} \geqslant 0$, $\sum_{k=1}^{n}\sum_{l=1}^{m} p_{kl} = 1$, $p_k = \sum_{l=1}^{m} p_{kl}$

atunci

$$H_n(p_1,\ldots,p_n) = -\lambda \sum_{k=1}^{n} p_k \log p_k \qquad (3.5.2)$$

unde λ este o constantă reală pozitivă.

Scanned with OKEN Scanner

3.5.3. Axiomatica lui D. K. Fadeev (1956)

Propoziția 3.5.3. Fie $\{H_n(p_1,\ldots,p_n)\}_{n\in\mathbf{N}}$ un șir de funcții nenegative definite pe domeniul $p_k \geqslant 0, 1 \leqslant k \leqslant n, \sum_{k=1}^{n} p_k = 1,$ $n \in \mathbf{N}$

Dacă

(F.1) $H_2(p, 1-p)$ este o funcție continuă în raport cu $p \in [0,1]$

(F.2) $H_n(p_1\ldots,p_n)$ este o funcție simetrică în raport cu toate argumentele pentru orice $n \in \mathbf{N}$

(F.3) $H_{n+1}(p_1,\ldots,p_{n-1}, p', p'') = H_n(p_1,\ldots,p_n) +$

$$+ p_n H_2\left(\frac{p'}{p_n}, \frac{p''}{p_n}\right)$$

unde $p_n + p' = p''$

atunci

$$H_n(p_1,\ldots,p_n) = -\lambda \sum_{k=1}^{n} p_k \log p_k \qquad (3.5.3)$$

unde λ este o constantă pozitivă.

Se observă cu ușurință asemănările și deosebirile dintre cele trei axiomatici, rămase pînă în prezent cele mai importante datorită în special caracterului lor profund intuitiv. În anii următori apariției lor le-au fost aduse unele modificări de formulare sau anumite simplificări în demonstrație.

Astfel:

1. în 1958 H. Tverberg înlocuiește axioma (F.1) cu condiția

(F.1′) $H_2(p, 1-p)$ este o funcție integrabilă Lebesgue pe $[0,1]$;

2. în 1959 A. Rényi dă o demonstrație simplificată axiomticii lui Fadeev;

3. în 1964 D.G. Kendall înlocuiește axioma (F.1) cu condiția (F.1″) $H_2(p, 1-p)$ este o funcție crescătoare pe $(0, 1/2]$;

4. în 1964 P.M. Lee înlocuiește și el axioma (F.1) cu condiția (F.1‴) $H_2(p, 1-p)$ este o funcție măsurabilă Lebesgue pe $(0,1)$, de fiecare dată celelalte axiome rămînînd aceleași.

Scanned with OKEN Scanner

3.5.4. Axiomatica lui H. Hatori (1958)

În 1958 H. Hatori a căutat să transpună axiomatica entropiei Shannon discrete la cazul entropiei continue

$$H(X) = -\int_{\mathbf{R}^n} p(x) \log p(x)\, \mathrm{d}x \tag{3.5.4}$$

unde $p(x)$ este densitatea de repartiție a vectorului aleator n-dimensional X, adică $p(x) \geqslant 0, \int_{\mathbf{R}^n} p(x)\mathrm{d}x = 1$.

Propoziția 3.5.4. Fie $X = (X_1, \ldots, X_n)$ un vector aleator continuu n-dimensional cu densitatea de repartiție $p(x) : D \subset \mathbf{R}^n \to [0, \infty)$ și entropia $H(X) = H(p)$. Dacă:

(H.1) entropia lui X, $H(p)$, este determinată numai de $p(x)$;

(H.2) $H(p) < H(q_D)$ pe $D \subset \mathbf{R}^n$, $p \neq q_D$, unde q_D este densitatea de repartiție uniformă pe $D, \int_D \mathrm{d}x < \infty$

$$q_D(x) = \begin{cases} \dfrac{1}{\int_D \mathrm{d}x}, & x \in D; \\ 0, & x \notin D \end{cases} \tag{3.5.5}$$

(H.3) $H(X) = H(Y) + H[(X - Y)/Y]$, unde Y este o parte a lui X iar $H[(X - Y)/Y]$ este entropia vectorului $(X - Y)$ condiționat de vectorul Y;

(H.4) dacă $p(x)$ ia un număr finit de valori $c_1, c_2, \ldots, c_m$ care satisfac relația $\sum_{k=1}^{m} c_k \mu_k = 1$, unde $\mu_k = \int_{A_k} \mathrm{d}x < \infty$,

$$A_k = \{x \in D \subseteq \mathbf{R}^n / p(x) = c_k\}, \quad 1 \leqslant k \leqslant m,$$

atunci $H(X)$ este funcție numai de constantele $c_1, c_2, \ldots, c_m$ și nu depinde de dimensiunea spațiului n;

atunci

$$H(X) = H(p) = -\lambda \int_D p(x) \log p(x)\mathrm{d}x \tag{3.5.6}$$

unde λ este o constantă pozitivă.

Scanned with OKEN Scanner

Să remarcăm că în ciuda străduaniei de transpunere a axiomaticii entropiei Shannon de la cazul discret la cazul continuu, proprietăţile acesteia din cazul discret nu pot fi prelungite la cazul continuu prin trecere la limită. Ideea lui Hatori rămîne însă şi trebuie apreciată ca atare. Ceea ce se desprinde din întregul efort de axiomatizare este ideea de unicitate a entropiei Shannon. Preocupări în direcţia axiomatizării entropiei continue au mai avut S. Ikeda (1962), L. L. Campbell (1965, 1972) şi alţii.

3.5.5. Epsilon-entropia lui A. N. Kolmogorov (1958)

În 1958 A.N. Kolmogorov a introdus entropia unui spaţiu metric, rezultat asupra căruia vor avea preocupări A. Rényi (1959), A.G. Vitushkin (1961) şi alţii.

Definiţia 3.5.1. Fie $\mathfrak{X}$ o mulţime arbitrară conţinînd n elemente. Numărul

$$K(\mathfrak{X}) = \log n \tag{3.5.7}$$

se numeşte entropie a mulţimii $\mathfrak{X}$.

Evident că dacă $\mathfrak{X}$ conţine o infinitate de elemente, atunci $K(\mathfrak{X}) = \infty$. Kolmogorov a sugerat o cale posibilă de ataşare a entropiei şi pentru mulţimi infinite, măsurîndule „masivitatea".

Fie $\{\mathfrak{X}, d\}$ un spaţiu metric cu definiţia obişnuită, total mărginit ceea ce înseamnă că pentru orice $\varepsilon > 0$ există o partiţie $\mathfrak{L}(\mathfrak{X})$ a lui $\mathfrak{X}$ astfel că diametrul fiecărui element al lui $\mathfrak{L}(\mathfrak{X})$ este inferior sau egal cu ε. Ca urmare $\mathfrak{L}(\mathfrak{X})$ este numită ε-partiţie a lui $\mathfrak{X}$. Fie $N_{\mathfrak{X}}(\varepsilon)$ numărul minimal al elementelor lui $\mathfrak{L}(\mathfrak{X})$ şi

$$A_1(\varepsilon), \ldots, A_{n(\varepsilon)}(\varepsilon)$$

o ε-partiţie a lui $\mathfrak{X}$. Presupunem că această ε-partiţie este asimptotic minimală în sensul că

$$\lim_{\varepsilon \to 0} \frac{\log n(\varepsilon)}{\log N_{\mathfrak{X}}(\varepsilon)} = 1$$

Fie acum X o variabilă aleatoare cu valori în $\mathfrak{X}$, definită pe cîmpul de probabilitate $\{\Omega, \mathfrak{K}, P\}$, avînd repartiţia

$$p_k(\varepsilon) \geqslant 0, \quad \sum_{k=1}^{n(\varepsilon)} p_k(\varepsilon) = 1 \text{ unde } p_k(\varepsilon) = P[X = x \in A_k(\varepsilon)]$$

Scanned with OKEN Scanner

Definiția 3.5.2. Mărimea $H(\varepsilon, X)$ dată de relația

$$H(\varepsilon, X) = -\sum_{k=1}^{n(\varepsilon)} p_k(\varepsilon) \log p_k(\varepsilon) \qquad (3.5.8.)$$

se numește ε-entropie a variabilei aleatoare X în raport cu partiția $\{A_k(\varepsilon)\}_{k=1}^{n(\varepsilon)}$ a spațiului metric $\mathfrak{X}$.

Analog entropiei Shannon clasice, se deduce imediat că

$$H(\varepsilon, X) \leqslant \log n(\varepsilon) \qquad (3.5.9)$$

Definiția 3.5.3. Punctele $\{x_1, x_2, \ldots, x_m\}$ din $\mathfrak{X}$ se zic ε-distinctibile dacă distanța dintre oricare două dintre ele depășește ε.

Să notăm cu $M_{\mathfrak{X}}(\varepsilon)$ numărul maxim de puncte ε-distinctibile ale lui $\mathfrak{X}$, număr care există și deci poate fi considerat ca o caracteristică a lui $\mathfrak{X}$.

Definiția 3.5.4. Dacă $\mathfrak{X}$ este un spațiu metric compact atunci se numește capacitatea acestuia numărul $C_{\mathfrak{X}}(\varepsilon)$ dat de relația

$$C_{\mathfrak{X}}(\varepsilon) = \log M_{\mathfrak{X}}(\varepsilon) \qquad (3.5.10)$$

În 1962 G.G. Lorentz a demonstrat că pentru orice spațiu metric compact $\mathfrak{X}$ și orice $\varepsilon > 0$, are loc relația

$$C_{\mathfrak{X}}(2\varepsilon) \leqslant \log N_{\mathfrak{X}}(\varepsilon) \leqslant C_{\mathfrak{X}}(\varepsilon) \qquad (3.5.11)$$

stabilind deci o mărginire a entropiei ε-partiției $\mathfrak{A}(\mathfrak{X})$ a lui $\mathfrak{X}$ cu ajutorul capacității lui $\mathfrak{X}$.

Spre deosebire de ε-entropia lui Kolmogorov, în 1967 E.C. Posner, E.R. Rodemich și H. Rumsey jr. au introdus o altă ε-entropie pe spații metrice probabilistice. Fie $\{\mathfrak{X}, d\}$ un spațiu metric și $\{\mathfrak{X}, \mathfrak{X}, P\}$ un spațiu probabilist în sensul clasic. O mulțime măsurabilă de diametru cel mult 2ε se numește pe scurt ε-mulțime sau ε-parte a lui $\mathfrak{X}$.

Definiția 3.5.5. Fiind dați $\varepsilon \geqslant 0$ și $\delta \geqslant 0$ se numește (ε, δ)-partiție a spațiului metric probabilist $\mathfrak{X}$ o partiție numărabilă $\mathfrak{A}(\mathfrak{X}) = \{A_i\}_{i \in N}$ a lui $\mathfrak{X}$, formată din ε-mulțimi disjuncte astfel că $P[\bigcup_{i \in N} A_i] = p \geqslant 1 - \delta$.

Dacă $\mathfrak{A}(\mathfrak{X})$ este o (ε, δ)-partiție a lui $\mathfrak{X}$, să notăm cu $p_i = P(A_i)$ și

$$q_i = \frac{p_i}{p} \qquad (3.5.12)$$

Scanned with OKEN Scanner

atunci expresia

$$H[\mathcal{Q}(\mathcal{X})] = -\sum_{i\in N} q_i \log\ q_i \qquad (3.5.13)$$

este definită ca entropie a repartiției $\{q_i\}_{i\in N}$ și $0 \leqslant H[\mathcal{Q}(\mathcal{X})] \leqslant \infty$.

Definiția 3.5.6. Se numește (ε, δ)-entropie a lui $\mathcal{X}$ mărimea

$$H_{\varepsilon,\delta}(\mathcal{X}) = \inf\{H[\mathcal{Q}(\mathcal{X})]\} \qquad (3.5.14)$$

inferiorul fiind luat după toate (ε, δ)-partițiile posibile ale lui $\mathcal{X}$ și ε-entropie a lui $\mathcal{X}$ mărimea

$$H_\varepsilon(\mathcal{X}) = H_{\varepsilon,0}(\mathcal{X}) \qquad (3.5.15)$$

Se demonstrează că dacă $\delta > 0$, atunci $H_{\varepsilon,\delta}(\mathcal{X})$ este finită (S. Guiașu, 1977). Deși într-un fel diferite, cele două ε-entropii dau același lucru și anume cantitatea minimă de informație necesară descrierii elementelor lui $\mathcal{X}$. Conexiunea dintre ele este similară conexiunii dintre entropia lui Hartley și cea a lui Shannon.

3.5.6. Axiomatica și alfa- entropia lui A. Rényi (1961)

Preocupat de entropia lui Shannon, A. Rényi a dat în 1961 o axiomatică privind definiția unei măsuri a nedeterminării în cazul unei repartiții probabilistice incomplete $\left(\sum_{k=1}^{n} p_k \leqslant 1\right)$, rezultat asupra căruia el va mai reveni ulterior și căruia i s-au adus diverse extensiuni.

Definiția 3.5.7. Fiind dată repartiția probabilistică incompletă $p_k > 0$, $\sum_{k=1}^{n} p_k \leqslant 1$, se numește α-entropie în sensul lui Rényi expresia

$$H_n^\alpha(p_1, \ldots, p_n) = \frac{1}{1-\alpha} \log \frac{\sum_{k=1}^{n} p_k^\alpha}{\sum_{k=1}^{n} p_k} \quad (\alpha > 0,\ \alpha \neq 1) \qquad (3.5.16)$$

Deducem imediat

Propoziția 3.5.5. Dacă $\alpha \to 1$ și $\sum_{k=1}^{n} p_k = 1$, entropiile Rényi și Shannon coincid, adică

$$\lim_{\alpha \to 1} H_n^{\alpha}(p_1, \ldots, p_n) = -\sum_{k=1}^{n} p_k \log_2 p_k = H_n(p_1, \ldots, p_n) \tag{3.5.17}$$

Propoziția 3.5.6. Fie $p_k > 0$, $\sum_{k=1}^{n} p_k \leqslant 1$, o repartiție probabilistică incompletă. Dacă

(R.1) $H_1\left(\frac{1}{2}\right) = 1$

(R.2) $H_{mn}(p_1q_1, \ldots, p_1q_n, \ldots, p_nq_1, \ldots, p_mq_n) = H_m(p_1, \ldots, p_m) + H_n(q_1, \ldots, q_n)$

(R.3) (axioma mediei cvasiaritmetice)

$$H_n(p_1, \ldots, p_n) = g^{-1}\left[\frac{\sum_{k=1}^{n} p_k g(H_1(p_k))}{\sum_{k=1}^{n} p_k}\right]$$

(R.4) H_1 este continuă

(R.5) g este crescătoare pe $[0, \infty)$

(R.6) g este continuă pe $[0, \infty)$

atunci există atît entropia lui Shannon incompletă

$$H_n(p_1, \ldots, p_n) = -\frac{\sum_{k=1}^{n} p_k \log p_k}{\sum_{k=1}^{n} p_k} \tag{3.5.18}$$

cît și α-entropia

$$H_n^{\alpha}(p_1, \ldots, p_n) = \frac{1}{1-\alpha} \log \frac{\sum_{k=1}^{n} p_k^{\alpha}}{\sum_{k=1}^{n} p_k} \quad (\alpha > 0, \alpha \neq 1) \tag{3.5.19}$$

Demonstația propoziției 3.5.6 a fost dată în 1963 de către Z. Daróczy. În același an, J. Aczél și Z. Daróczy au demonstrat că axioma (R.2) împreună cu următoarele trei axiome :

(A-D.1) dacă $\sum_{k=1}^{n} p_k = 1$, atunci

$$H_n(p_1, \ldots, p_n) = g^{-1}\left[\sum_{k=1}^{n} p_k\, g(-\log p_k)\right]; \tag{3.5.20}$$

Scanned with OKEN Scanner

(A-D.2) funcția

$$f(x) = \begin{cases} x\, g(-\log x) & x > 0 \\ 0 & x = 0 \end{cases} \tag{3.5.21}$$

este strict convexă pe $[0, 1]$;

(A-D.3) funcția g este continuă pe $[0, \infty)$ implică existența entropiei Shannon

$$H_n(p_1, \ldots, p_n) = -\sum_{k=1}^{n} p_k \log p_k \tag{3.5.22}$$

cît și a α-entropiei Rényi

$$H_n^{\alpha}(p_1, \ldots, p_n) = \frac{1}{1-\alpha} \log \left(\sum_{k=1}^{n} p_k^{\alpha} \right) \tag{3.5.23}$$

pentru o repartiție completă.

În 1964 Z. Daróczy revine și înlocuiește axioma (A-D.2) prin și (R.5) (D.1) $\lim_{x \to \infty} [x g(-\log x] = 0$.

După apariția acestor rezultate, diferite extensii ale α-entropiei Rényi au fost propuse în anii următori.

(E.1) În anul 1966 R. S. Varma a propus două noi expresiale lui H_n^{α} și anume:

$$H_n^{\alpha m}(p_1, \ldots, p_n) = \frac{1}{m-\alpha} \log \frac{\sum_{k=1}^{n} p_k^{\alpha-m+1}}{\sum_{k=1}^{n} p_k},$$

$$m - 1 < \alpha < m,\ m \geqslant 1,\ \sum_{k=1}^{n} p_k \leqslant 1 \tag{3.5.24}$$

și

$$H_n^{\alpha m}(p_1, \ldots, p_n) = \frac{1}{m(m-\alpha)} \log \frac{\sum_{k=1}^{n} p_k^{\alpha/m}}{\sum_{k=1}^{n} p},$$

$$0 < \alpha < m,\ m \geqslant 1,\ \sum_{k=1}^{n} p_k \leqslant 1 \tag{3.5.25}$$

Scanned with OKEN Scanner

ambele expresii coincizînd cu entropia lui Rényi pentru $m = 1$, dar în cazul particular $0 < \alpha < 1$ al relației (3.5.16).

(E.2) În anul 1967, J. N. Kapur a propus așa-numita entropie de ordin α și tip β, definită de expresia

$$H_n^{\alpha\beta}(p_1, \ldots, p_n) = \frac{1}{1-\alpha} \log \frac{\sum_{k=1}^{n} p_k^{\alpha+\beta-1}}{\sum_{k=1}^{n} p_k^{\beta}},$$

$$\alpha > 0,\ \alpha \neq 1,\ \beta \geqslant 1,\ \sum_{k=1}^{n} p_k \leqslant 1 \qquad (3.5.26)$$

care coincide cu α-entropia lui Rényi dacă $\beta = 1$.

(E.3) În anul 1970, P. N. Rathie generalizează rezultatul lui J. N. Kapur propunînd entropia de ordin α și tip (β_k) $1 \leqslant k \leqslant n$, a cărei expresie este

$$H_n^{\alpha,\beta_1,\ldots,\beta_n}(p_1, \ldots, p_n) = \frac{1}{1-\alpha} \log \frac{\sum_{k=1}^{n} p_k^{\alpha+\beta_k-1}}{\sum_{n=1}^{n} p_k^{\beta_k}},$$

$$\alpha > 0,\ \alpha \neq 1,\ \beta_k \geqslant 1,\ \sum_{k=1}^{n} p_k \leqslant 1 \qquad (3.5.27)$$

și care pentru $\beta_k = \beta$, $1 \leqslant k \leqslant n$, coincide evident cu (3.5.26).

Să remarcăm că din extensiile de mai sus putem deduce imediat o altă generalizare și anume

$$H_n^{\alpha,m,\beta_1,\ldots,\beta_n}(p_1, \ldots, p_n) = \frac{1}{m-\alpha} \log \frac{\sum_{k=1}^{n} p_k^{\alpha+\beta_k-m}}{\sum_{k=1}^{n} p_k^{\beta_k}},$$

$$\alpha > m-1,\ \alpha \neq m,\ \beta_k \geqslant 1,\ \sum_{k=1}^{n} p_k \leqslant 1 \qquad (3.5.28)$$

care, pentru $m = 1$ coincide cu (3.5.27), pentru $\beta_k = 1$, $1 \leqslant k \leqslant n$, constituind o extensie a lui (3.5.24) la cazul $\alpha > m-1$, $\alpha \neq m$, iar pentru $m = \beta_k = 1$, $1 \leqslant k \leqslant n$ coincide cu α-entropia lui Rényi. Deducem imediat

Scanned with OKEN Scanner

Propoziția 3.5.7. Pentru orice repartiție $p_k>0, \sum_{k=1}^{n} p_k \leqslant 1$, avem

$$H_n^0(p_1, \ldots, p_n) \geqslant \log_2 n \tag{3.5.29}$$

egalitatea dintre entropia lui Rényi de ordin zero și entropia Hartley avînd loc dacă și numai dacă repartiția este completă.

Definiția 3.5.8. Fiind dată repartiția $p_k \geqslant 0, \sum_{k=1}^{n} p_k = 1$, expresia

$$E_n(p_1, \ldots, p_n) = \sum_{k=1}^{n} p_k^2 \tag{3.5.30}$$

se numește energie informațională Onicescu.

Următorul rezultat este de asemenea imediat.

Propoziția 3.5.8. Pentru orice repartiție completă $p_k \geqslant 0, \sum_{k=1}^{n} p_k = 1$, între energia informațională Onicescu și entropia lui Rényi are loc relația

$$H_n^2[p_1, \ldots, p_n] = -\log E_n[p_1, \ldots, p_n] \text{ sau } E_n[p_1, \ldots, p_n] = 2^{-H_n^2[p_1, \ldots, p_n]} \tag{3.5.31}$$

Ca urmare a propoziției 3.5.8. rezultă că în cazul $\alpha = 2$, α-entropia Rényi și energia informațională Onicescu se determină reciproc și în mod unic.

3.5.7. Axiomatica lui R. S. Ingarden și K. Urbanik (1962)

Axiomaticile prezentate privind determinarea entropiei Shannon au la bază ideea definirii măsurii cantității de informație cu ajutorul repartiției probabilistice, presupusă cunoscută, ca urmare firească a interdependenței dintre conceptele de informație și probabilitate. Deoarece în practică avem în general diferite informații despre un anumit experiment aleator, fără a cunoaște însă repartiția probabilistică a acestuia, s-a pus problema determinării probabilității pe baza informației presupusă cunoscută și măsurabilă. Prima încercare de soluționare a acestei probleme aparține lui R.S. Ingarden și K. Urbanik, care au propus în 1962 un

sistem de axiome pentru determinarea entropiei care nu se bazează pe repartiţia probabilistică a experimentului în studiu şi ca urmare, au obţinut din aceste axiome atît existenţa repartiţiei de probabilităţi cît şi expresia entropiei Shannon în funcţie de aceste probabilităţi, aceeaşi ca şi în axiomaticile precedente. Amănunte asupra acestei axiomatici, care fiind mai greoaie nu o dăm aici, se găsesc în diferite lucrări ce abordează această problemă (Guiaşu (1966, 1977). Dezavantajul axiomaticii lui Ingarden şi Urbanik se datorează îndeosebi formei foarte complicate a axiomelor cerute noţiunii de informaţie, ceea ce le-a făcut greu acceptabile din punct de vedere intuitiv, ca propoziţii de pornire în construcţia teoriei informaţiei şi a teoriei probabilităţilor pe baza conceptului de informaţie. Cu toate acestea, ideile lui Ingarden şi Urbanik au însă o anumită importanţă teoretică deoarece constituie prima încercare de a determina probabilitatea cu ajutorul informaţiei. Ingarden în 1965 şi 1977 şi Urbanik în 1973 au adus unele simplificări asupra teoriei lor, dar cu toate acestea ea rămîne încă greoaie, poate tocmai pentru generalitatea sa în prezentarea entropiei Shannon. Rezultatele lor, sub titlul „informaţie fără probabilitate", au constituit subiect de reflecţie pentru mulţi din cei preocupaţi de teoria informaţiei.

3.5.8. Entropia Onicescu (1963)

Pornind de la entropia clasică a lui Shannon, O. Onicescu, a dat în 1963 o generalizare a acesteia prin introducerea conceptului de entropie a unui eveniment relativă la o partiţie a evenimentului sigur.

Fie aşadar $\{\Omega, \mathcal{K}, P\}$ un cîmp de probabilitate, $\mathcal{A} = \{A_j\}_{j=1}^m, A_j \cap A_k = \emptyset, j \neq k, \bigcup_{j=1}^m A_j = \Omega$ o partiţie a lui Ω şi $C \in \mathcal{K}$ un eveniment oarecare.

Definiţia 3.5.9. Se numeşte entropie a evenimentului C relativă la partiţia $\mathcal{A}$ în sensul lui Onicescu, expresia

$$H_{\mathcal{A}}(C) = -\sum_{j=1}^{m} P(A_j \cap C) \log_2 P(A_j) \qquad (3.5.32)$$

Propoziţia 3.5.9. Entropia Onicescu satisface următoarele

1) dacă $C = \Omega$ atunci entropiile Onicescu şi Shannon coincid

Scanned with OKEN Scanner

$$H_{\mathcal{A}}(\Omega) = -\sum_{j=1}^{m} P(A_j) \log P(A_j) \qquad (3.5.33)$$

2) dacă C este independent de $\mathcal{A}$ atunci

$$H_{\mathcal{A}}(C) = P(C)\, H_{\mathcal{A}}(\Omega) \qquad (3.5.34)$$

3) dacă $C = \bigcup_{k=1}^{n} C_k$, $C_k \cap C_l = \emptyset$, $k \neq l$, atunci

$$H_{\mathcal{A}}(C) = \sum_{k=1}^{n} H_{\mathcal{A}}(C_k) \qquad (3.5.35)$$

4) dacă $\mathcal{A} = \{A_j\}_{j=1}^{m}$ și $\mathcal{B} = \{B_k\}_{k=1}^{n}$ sînt două partiții arbitrare ale lui Ω atunci

$$H_{\mathcal{A}\cap\mathcal{B}}(C) = H_{\mathcal{A}}(C) + H_{\mathcal{B}/\mathcal{A}}(C) = H_{\mathcal{B}}(C) + H_{\mathcal{A}/\mathcal{B}}(C) \qquad (3.5.36)$$

unde

$$H_{\mathcal{B}/\mathcal{A}}(C) = -\sum_{j=1}^{m}\sum_{k=1}^{n} P(A_j \cap B_k \cap C) \log(B_k/A_j) \qquad (3.5.37)$$

5) dacă $\mathcal{A}$ și $\mathcal{B}$ sînt independente, atunci

$$H_{\mathcal{A}\cap\mathcal{B}}(C) = H_{\mathcal{A}}(C) + H_{\mathcal{B}}(C) \qquad (3.5.38)$$

6) $H_{\mathcal{A}}(C) \geqslant 0$, cu egalitate dacă și numai dacă $\mathcal{A} = \{\Omega, \emptyset\}$

7) este simetrică în raport cu evenimentele partiției

8) $$H_{\mathcal{A}}(C) \leqslant H_{\mathcal{A}}\Omega) \leqslant \log m \qquad (3.5.39)$$

cu prima egalitate dacă și numai dacă $C = \Omega$, cu a doua dacă și numai dacă evenimentele partiției $\mathcal{A}$ sînt echiprobabile și cu amîndouă dacă și numai dacă $C = \Omega$ și $\mathcal{A}$ are elementele egal probabile.

Demonstrația este imediată. Să remarcăm că entropia Onicescu poate fi definită și pentru vectori aleatori.

3.5.9. Entropia Rényi d-dimensională de ordin α (1970)

În 1970 A. Rényi revine asupra unor rezultate din anii precedenți și ținînd cont de epsilon-entropia lui Kolmogorov și de α-entropia sa, introduce conceptul de entropie d-dimensională de ordin α.

Scanned with OKEN Scanner

Fie $p = \{p_k\}_{k\in\mathbb{N}}$ și $q = \{q_k\}_{k\in\mathbb{N}}$ repartițiile variabilelor aleatoare discrete numărabile X și Y.

Dacă

$$H^\alpha(X) = \frac{1}{1-\alpha} \log \sum_{k=1}^{\infty} p_k^\alpha, \quad \alpha \geqslant 0, \ \alpha \neq 1 \tag{3.5.40}$$

reprezintă α-entropia lui X sau informația de ordin α conținută în X, unde

$$H^1(X) = H(X) = -\sum_{k=1}^{\infty} p_k \log p_k \tag{3.5.41}$$

cînd cele două serii sînt convergente, atunci putem introduce variația de informație, sau cîștigul de informație, de ordin α cînd se trece de la repartiția q la repartiția p, ca fiind

$$H^\alpha(p/q) = \frac{1}{1-\alpha} \log_2 \left(\sum_{k=1}^{\infty} \frac{p_k^\alpha}{q_k^{\alpha-1}} \right), \quad \alpha \neq 1 \tag{3.5.42}$$

și

$$H^1(p/q) = H(p/q) = \sum_{k=1}^{\infty} p_k \log \frac{p_k}{q_k} \tag{3.5.43}$$

cînd de asemenea cele două serii sînt convergente. Ultima relație reprezintă variația medie de informație cînd se trece de la repartiția q la repartiția p, avînd variația individuală

$$(-\log q_k - (-\log p_k)) = \log \frac{p_k}{q_k}.$$

Să presupunem că X este o variabilă continuă și să considerăm variabila $X_n = \dfrac{[nX]}{n}$, unde $[a]$ este partea întreagă a lui a. Să presupunem de asemenea $\alpha > 0$ și că $H^\alpha(X_n) < \infty$ pentru orice n, precum și faptul că limita

$$d = d_\alpha(X) = \lim_{n\to\infty} \frac{H^\alpha(X_n)}{\log n} \tag{3.5.44}$$

există.

Definiția 3.5.10. Cantitatea $d_\alpha(X)$ dată de relația (3.5.44) se numește dimensiunea de ordin α a lui X, iar limita

$$H^{\alpha,d}(X) = \lim_{n\to\infty} (H^\alpha(X_n) - d \log n) \tag{3.5.45}$$

Scanned with OKEN Scanner

dacă există, se numește informație d-dimensională de ordin α conținută în X, sau entropie d-dimensională de ordin α a lui X.

În baza acestor concepte Rényi a dedus

Propoziția 3.5.10. Fie X o variabilă aleatoare continuă cu densitatea $p(x)$ și $X_n = \frac{[nX]}{n}$, $n \in \mathbf{N}$.

a) dacă $H^\alpha(X_1) < \infty$, $\alpha > 0$, atunci $\lim\limits_{n\to\infty} \frac{H^\alpha(X_n)}{\log n} = 1$

b) dacă integrala $\int\limits_R (p(x))^\alpha \, dx$, $\alpha \neq 1$ există, atunci

$$\lim_{n\to\infty}(H^\alpha(X_n) - \log n) = H^{\alpha,1}(X) = \frac{1}{1-\alpha} \log\left(\int_R (p(x))^\alpha \, dx\right) \tag{3.5.46}$$

c) dacă integrala $-\int\limits_R p(x) \log p(x) \, dx$, există, atunci

$$H^{1,1}(X) = \lim_{n\to\infty}(H^1(X_n) - \log n) = H(X) = -$$

$$-\int_R p(x) \log p(x) \, dx \tag{3.5.47}$$

Să remarcăm că propoziția 3.5.10 nu este altceva decît un mod de a deduce entropia lui Boltzmann sau a lui Shannon pentru o repartiție continuă, particularizînd α-entropia lui Rényi pentru cazul continuu, ea însăși un caz particular al entropiei d-dimensionale de ordin α.

3.5.10. Alfa-entropia Behara-Nath (1973)

M. Behara și P. Nath au definit în 1973, în alt mod decît Rényi, o măsură a cantității de informație care în particular coincide cu α-entropia lui Rényi precum și cu entropia Shannon.

Scanned with OKEN Scanner

Fie funcția reală

$$Z_\alpha(t) = \begin{cases} \dfrac{t - t^\alpha}{1 - 2^{1-\alpha}} & t \in (0, 1] \text{ și } \alpha \in [0, \infty) \\ 1 & t = \alpha = 0 \\ 0 & t = 0 \text{ și } \alpha \in (0, \infty) \end{cases} \tag{3.5.48}$$

Deoarece $\lim_{\alpha \to 1} Z_\alpha(t) = -t \log t$, rezultă că

$$Z_1(t) = \begin{cases} -t \log t & t \in (0, 1] \\ 0 & t = 0 \end{cases} \tag{3.5.49}$$

Deducem imediat

Propoziția 3.5.11. Pentru funcția $Z_\alpha(t)$ sînt adevărate afirmațiile :

1) $Z_\alpha(t) \geqslant 0$ pentru $t \in [0,1]$ și $\alpha \in [0, \infty)$

2) $Z_\alpha(t)$ este continuă în raport cu α și t.

3) $Z_\alpha(t)$, $\alpha > 0$, $\alpha \neq 1$, este strict crescătoare în raport cu t pe $[0, \alpha^{\frac{1}{1-\alpha}}]$ și strict descrescătoare pe $[\alpha^{\frac{1}{1-\alpha}}, 1]$ admițînd

$$[\max_t] Z_\alpha(t) = \frac{(\alpha - 1)\alpha^{\frac{\alpha}{1-\alpha}}}{1 - 2^{1-\alpha}} = Z_\alpha(\alpha^{\frac{1}{1-\alpha}}), \quad \alpha > 0, \quad \alpha \neq 1$$

4) $Z_0(t)$ este strict descrescătoare pe $[0,1]$ cu $Z_0(0) = 1$ și $Z_0(1) = 0$

5) $Z_1(t)$ crește strict pe $[0, 1/2]$ și descrește strict pe $[1/2, 1]$, cu $Z_1(0) = Z_1(1) = 0$ și $Z_1(1/2) = 1/2$.

Fie acum $\{\Omega, \mathfrak{K}, P\}$ un cîmp de probabilitate și $\mathcal{A}$ o partiție finită a lui Ω.

Definiția 3.5.11. Se numește α-entropie a partiției finite $\mathcal{A}$ în sensul lui Behara și Nath, expresia

$$I^\alpha(\mathcal{A}) = \sum_{A \in \mathcal{A}} Z_\alpha(P(A)) = \frac{1}{1 - 2^{1-\alpha}} \left[1 - \sum_{A \in \mathcal{A}} P^\alpha(A)\right], \quad \alpha \geqslant 0, \ \alpha \neq 1 \tag{3.5.50}$$

Deducem imediat

Propoziția 3.5.12. Avem

$$\lim_{\alpha \to 1} I^\alpha(\mathcal{A}) = \lim_{\alpha \to 1} H^\alpha(\mathcal{A}) = -\sum_{A \in \mathcal{A}} P(A) \log_2 P(A) = H(\mathcal{A}) \tag{3.5.51}$$

Scanned with OKEN Scanner

Observăm că relația (3.5.51) justifică faptul că (3.5.50) este o extensiune a entropiei Shannon, $H(\mathcal{A})$ fiind o altă formă de exprimare a relațiilor (3.1.8) și (3.1.9).

Să considerăm acum partiția $\mathcal{A} = \{A_1, A_2\}$, $A_1 \neq \emptyset$, $A_2 \neq \emptyset$ $A_1 \cup A_2 = \Omega$, $P(A_1) = p \in (0,1)$. În acest caz avem

$$I^\alpha(\mathcal{A}) = Z_\alpha(p) + Z_\alpha(1-p) = \Phi_\alpha(p), \quad \alpha \geqslant 0 \tag{3.5.52}$$

și deducem fără dificultăți

Propoziția 3.5.13. Avem

1) $\Phi_\alpha(p) = \Phi_\alpha(1-p)$, $p \in (0,1)$, $\Phi_\alpha(1/2) = 1$

2) $\Phi_0(p) = 1$, $\Phi_1(p) = H_2[p, 1-p]$; $\Phi_2(p) = \Phi_3(p) = 4p(1-p)$

3) $\Phi_\alpha(p) \neq \Phi_{\alpha-1}(p)$ pentru $\alpha = 4, 5, 6, \ldots$ și $p \in (0,1)$

4) $I^0(\mathcal{A}) = 1$; $I^1(\mathcal{A}) = H_2[p, 1-p]$; $I^2(\mathcal{A}) = I^3(\mathcal{A}) = 4p(1-p)$

Pornind de la faptul că $Z_2(t) = 2(t - t^2)$ este ecuația unei parabole, Behara și Nath au denumit expresia $I^2(\mathcal{A})$ entropie parabolică a partiției $\mathcal{A}$. Deducem cu ușurință

Propoziția 3.5.14. Pentru orice partiție finită $\mathcal{A}$ a lui Ω între α-entropia Behara-Nath și α-entropia Rényi are loc relația

$$I^\alpha(\mathcal{A}) = \frac{1}{1 - 2^{1-\alpha}} \left[1 - 2^{(1-\alpha)H^\alpha(\mathcal{A})}\right]$$

sau

$$H^\alpha(\mathcal{A}) = \frac{1}{1-\alpha} \log\left[1 - (1 - 2^{1-\alpha})\, I^\alpha(\mathcal{A})\right] \tag{3.5.53}$$

Relația (3.5.53) ne arată că deși aparent sînt total diferite, entropiile Rényi și Behara-Nath se determină în mod reciproc și unic. Deducem de asemenea imediat

Propoziția 3.5.15. Între entropia parabolică Behara-Nath și energia informațională Onicescu avem

$$I^2(\mathcal{A}) = 2[1 - E(\mathcal{A})] \text{ sau } E(\mathcal{A}) = 1 - \frac{1}{2} I^2(\mathcal{A}) \tag{3.5.54}$$

pentru orice partiție $\mathcal{A}$ a lui Ω. Relația (3.5.54) ne sugerează

Scanned with OKEN Scanner

Definiția 3.5.12. Se numește energie informațională Onicescu generalizată de ordin α atașată partiției finite $\mathcal{A}$ a lui Ω expresia

$$E^{\alpha}(\mathcal{A}) = \sum_{A \in \mathcal{A}} P^{\alpha}(A), \quad \alpha \geqslant 0, \quad \alpha \neq 1 \tag{3.5.55}$$

Ca urmare deducem

Propoziția 3.5.16. Între alfa-entropia Behara-Nath și energia informațională Onicescu generalizată are loc relația

$$I^{\alpha}(\mathcal{A}) = \frac{1}{1 - 2^{1-\alpha}} [1 - E^{\alpha}(\mathcal{A})]. \tag{3.5.56}$$

Dacă în locul partiției finite $\mathcal{A}$ considerăm o variabilă aleatoare discretă finită X a cărei repartiție este $p_k \geqslant 0$, $\sum_{k=1}^{n} p_k = 1$, atunci

$$I^{\alpha}(X) = \frac{1}{1 - 2^{1-\alpha}} \left[1 - \sum_{k=1}^{n} p_k^{\alpha}\right] = I_n^{\alpha}[p_1, \ldots, p_n] \tag{3.5.57}$$

și

$$E^{\alpha}(X) = \sum_{k=1}^{n} p_k^{\alpha} = E_n^{\alpha}(p_1, \ldots, p_n) \tag{3.5.58}$$

Observăm cu ușurință că

a) dacă $\alpha < 1$, atunci $1 \leqslant E^{\alpha}(X) \leqslant n^{1-\alpha}$, egalitatea a doua fiind atinsă dacă și numai dacă avem o repartiție uniformă discretă

b) dacă $\alpha \geqslant 1$, atunci $n^{1-\alpha} \leqslant E^{\alpha}(X) \leqslant 1$, prima egalitate avînd loc în cazul repartiției uniforme, iar cea de a doua pentru $\alpha = 1$ sau dacă $p_k = 1$ și $p_j = 0$, $j \neq k$, $1 \leqslant j \leqslant n$, deci cînd X este o constantă.

Remarcăm că alfa-entropia Behara-Nath poate fi prezentată și pentru o partiție numărabilă $\mathcal{A}$ a lui Ω cu condiția ca seria corespunzătoare să fie convergentă, condiție de altfel îndeplinită dacă $\alpha > 1$.

Să observăm de asemenea că alfa-entropia Behara-Nath poate fi definită și pentru repartiții continue și anume

$$I^{\alpha}(X) = \frac{1}{1 - 2^{1-\alpha}} \left[1 - \int_{R} p^{\alpha}(x)\, dx\right] \tag{3.5.59}$$

Scanned with OKEN Scanner

cu condiția ca funcția $p^{\alpha}(x)$ să fie integrabilă, de unde pentru $\alpha \to 1$ obținem

$$I^1(X) = -\int_R p(x) \log p(x)\, dx \qquad (3.5.60)$$

relație care ne arată că pentru $\alpha = 1$ entropiile Behara-Nath și Shannon coincid și în cazul repartițiilor continue.

Axiomaticile și extensiunile entropiei Shannon prezentate mai sus, sînt ideile și rezultatele cele mai importante formulate de la apariția teoriei informației asupra măsurii cantității de informație, dar nu și singurele. Important rămîne însă un fapt și anume acela că toate au un scop comun, de a dovedi că informația este măsurabilă cantitativ și că măsura propusă de Shannon rămîne la baza teoriei matematice a informației.

4. O măsură a utilității informației

4.1. Utilitate

Desigur, definirea unei măsuri a cantității de informație pe care o conține și deci o furnizează un anumit experiment, este un rezultat remarcabil și constituie fără îndoială primul pas fundamental al elaborării teoriei matematice a informației. Dar și aici, ca și în multe alte domenii ale științei și nu numai ale ei, rezolvarea unei probleme pune o altă problemă. Astfel, apare în mod natural întrebarea :„două rezultate ale unui experiment aleator, avînd șanse egale sau diferite de realizare, furnizează cantități de informație la fel de importante sau la fel de utile ? " Evident că întrebarea poate fi pusă în cazul general al unui experiment cu mai multe rezultate posibile, egal probabile sau nu. O astfel de problemă a condus în mod firesc la studierea posibilității de măsurare a utilității informației furnizate de un anumit experiment aleator. Conceptul de utilitate a informației, în sensul său cel mai larg, a apărut pentru prima dată pe la 1730 și aparține cunoscutului matematician elvețian Gabriel Cramer (1704—1752). Noțiunea a fost consacrată însă, cu precădere în direcția teoriei jocurilor printr-o lucrare a matematicianului elvețian de renume Daniel Bernoulli (1700—1782), publicată în 1730 și va fi analizată riguros după mai

Scanned with OKEN Scanner

bine de două secole, cînd teoria probabilităţilor va fi şi ea o teorie matematică riguros fundamentată. Primii care s-au ocupat matematic de conceptul de utiliate, considerată ca o cuantificare a preferinţelor, pentru care au formulat primul sistem de axiome, au fost americanii John von Neumann (1903—1957), matematician de renume mondial şi Oskar Morgenstern (1902—1977), economist abil în utilizarea aparatului matematic, într-o lucrare din 1944 şi reeditată în 1947, care fundamenta teoria matematică a jocurilor. Axiomatica lor a stîrnit vii controverse şi a condus la formularea altor axiomatici precum cea dată de R.D. Luce şi H. Raiffa în 1957 şi altele, al căror rezultat final s-a concretizat de fapt în studierea riguros matematică a funcţiei utilitate sau a utilităţii corespunzătoare unui anumit experiment aleator. În linii mari, utilitatea a fos concepută ca o măsură a preferinţei faţă de un rezultat sau altul al unui experiment oarecare, sau faţă de un experiment.

Să considerăm un cîmp de probabilitate finit $\{\Omega, \mathcal{K}, P\}$ care poate descrie un experiment aleator cu un număr finit de rezultate. Pe $\mathcal{K}$ să definim o relaţie de preferinţă, notînd $A_1 \succcurlyeq A_2$ şi citind „A_1 este preferat lui A_2" sau „A_2 nu este preferat lui A_1". Dacă $A_1 \succcurlyeq A_2$ şi $A_2 \succcurlyeq A_1$, atunci scriem $A_1 \sim A_2$ şi citim „A_1 şi A_2 sînt egal preferabile, sau egal nepreferabile". Dacă $P(A_1) = p_1$ şi $P(A_2) = 1 - p_1$, atunci spunem că A_1 şi A_2 constituie o alternativă şi o notăm cu $[p_1, A_1; 1 - p_1, A_2]$. Noţiunea poate fi extinsă la cazul a n rezultate incompatibile două cîte două, scriind o astfel de alternativă sau experiment ca $[p_1, A_1; p_2, A_2; \ldots; p_n, A_n]$, $p_k \geqslant 0$, $\sum_{k=1}^{n} p_k = 1$.

4.1.1. Axiomatica Von Neumann – Morgenstern (1944)

(N-M.1) Pentru orice două evenimente A_i şi $A_j \in \mathcal{K}$ una şi numai una din următoarele situaţii are loc: $A_i \succ A_j$, $A_j \succ A_i$, $A_i \sim A_j$
Aceasta înseamnă că oricare două evenimente ale lui $\mathcal{K}$ sînt direct comparabile şi ca urmare sau unul este preferat celuilalt, sau cele două sînt egal preferate.

(N-M.2) Dacă $A_i \succ A_j$ şi $A_j \succ A_k$, atunci $A_i \succ A_k$. Aceasta înseamnă că relaţia de preferinţă ($\succ$) este tranzitivă.

(N-M.3) Dacă $A_i \succ A_j$ atunci $A_i \succ [p, A_i; 1 - p, A_j]$ pentru orice $p \in (0, 1)$

Scanned with OKEN Scanner

Altfel spus, dacă A_i este preferat lui A_j, atunci A_i este preferat oricărei alternative ce cuprinde pe A_i și pe A_j.

(N-M.4) Dacă $A_j \succ A_i$ atunci $[p, A_i; 1-p, A_j] \succ A_i$ pentru orice $p \in (0,1)$.
Aceasta este duala axiomei (N-M.3) și înseamnă că dacă A_j este preferat lui A_i, atunci orice alternativă ce cuprinde A_i și A_j este preferată lui A_i.

(N-M.5) Dacă $A_i \succ A_j \succ A_k$, atunci există $p \in (0,1)$ astfel ca $[p, A_i; 1-p; A_k] \succ A_j$.
Ca urmare, dacă A_i este preferat lui A_j și A_j este preferat lui A_k, atunci există o alternativă cuprinzînd A_i și A_k, preferată lui A_j.

(N-M.6) Dacă $A_k \succ A_j \succ A_i$, atunci există $p \in (0,1)$ astfel încît

$$A_j \succ [p, A_i; 1-p, A_k].$$

Aceasta este duala axiomei (N-M.5). Ele pot fi formulate de altfel simultan, luînd o dată $p \in (0,1)$ și altă dată $q \in (0,1)$ și la fel pentru (N-M.3) și (N-M.4).

(N-M.7) $[p, A_i; 1-p, A_j] \sim [1-p, A_j; p, A_i]$.

Această echivalență înseamnă că ordinea rezultatelor într-o alternativă nu are nici o importanță.

(N-M.8) $$[p, (q, A_i; 1-q, A_j); 1-p, A_j] \equiv$$
$$\equiv [pq, A_i; 1-pq, A_j], \quad 0 \leqslant q \leqslant 1.$$

Identitatea înseamnă că alternativele compuse se pot descompune în alternative simple, folosind calculul adecvat al probabilităților, fără ca preferințele să fie afectate.

(N-M.9) Dacă $A_i \sim A_j$ atunci avem echivalența
$[p, A_i; 1-p, B] \sim [p, A_j; 1-p, B]$ pentru orice $p \in (0,1)$ și $B \in \mathcal{K}$.
Ca urmare, dacă $A_i \sim A_j$, atunci ele se pot înlocui reciproc într-o alternativă pe care o formează împreună cu un alt eveniment arbitrar B.

Definiția 4.1.1. Se numește utilitate în sensul axiomaticii von Neumann-Morgenstern orice funcție $u : \mathcal{K} \to \mathbf{R}$ care are proprietățile

(u.1) $A \succ B$ dacă și numai dacă $u(A) > u(B)$

(u.2) $u([p, A; 1-p, B]) = pu(A) + (1-p)u(B)$

Scanned with OKEN Scanner

Să remarcăm că (u.1) spune că utilitatea este monoton crescătoare în raport cu preferințele iar (u.2) spune că utilitatea unei alternative este egală cu valoarea medie a utilităților rezultatelor posibile ale alternativei.

Proprietatea (u.2) poate fi extinsă pentru un experiment cu n rezultate

$$[p_1, A_1; \ldots : p_n, A_n], \; p_k \geqslant 0, \; \sum_{k=1}^{n} p_k = 1$$

avînd forma

$$(u.2') \qquad u([p_1, A_1; \ldots; p_n, A_n]) = \sum_{k=1}^{n} p_k u(A_k)$$

Ca urmare a axiomaticii von Neumann-Morgenstern poate fi dedusă

Propoziția 4.1.1. Dacă u_1 este utilitate pe $\mathfrak{K}$ atunci $u_2 = au_1 + b$, $a > 0$, este de asemenea utilitate pe $\mathfrak{K}$.

4.1.2. Axiomatica Luce-Raiffa (1957)

(L-R.1) Pentru orice A_i și $A_j \in \mathfrak{K}$ avem sau $A_i \succcurlyeq A_j$ sau $A_j \succcurlyeq A_i$ iar dacă $A_i \succcurlyeq A_j$ și $A_j \succcurlyeq A_k$ atunci $A_i \succcurlyeq A_k$. Conform acestei axiome, a ordinii rezultatelor posibile, relația de preferință are loc pentru orice două rezultate posibile ale unui experiment și este tranzitivă.

Să presupunem acum, fără a diminua generalitatea problemei, că $A_1 \succcurlyeq A_2 \succcurlyeq \ldots \succcurlyeq A_n$ și că $A_1 \succ A_n$. (A_1 este strict preferat lui A_n), $A_i \cap A_j = \emptyset$, $\bigcup_{i=1}^{n} A_i = \Omega$.

Să notăm cu E experimentul $[p_1, A_1; p_2, A_2; \ldots; p_n, A_n]$. Fie atunci $E_1, E_2, \ldots, E_s$ s experimente, fiecare din ele avînd rezultatele $A_1, A_2, \ldots, A_n$. Dacă $q_k \geqslant 0$, $\sum_{k=1}^{s} q_k = 1$, atunci $[q_1, E_1; \ldots; q_s, E_s]$ formează un experiment compus.

(L-R.2) Un experiment compus este echivalent cu un experiment simplu și anume:

dacă $E_i = [p_1^{(i)}, A_1, \ldots, p_n^{(i)}, A_n]$, $1 \leqslant i \leqslant s$

atunci $[q_1, E_1; \ldots; q_s, E_s] \sim [p_1 A_1; \ldots; p_n, A_n]$

unde $p_k = q_1 p_k^{(1)} + q_2 p_k^{(2)} + \ldots + q_s p_k^{(s)}$, $1 \leqslant k \leqslant n$

Această axiomă, a reducerii experimentelor compuse, exprimă echivalența dintre un experiment compus și un experiment simplu, ale cărui rezultate sînt descrise de evenimentele elementare ale cîmpului.

Scanned with OKEN Scanner

(L—R.3) Orice eveniment elementar $A_i \in \mathcal{K}$ este echivalent cu experimentul ale cărui rezultate sînt doar A_1 și A_n. Adică există $r_i \in (0, 1)$ astfel încît $A_i \sim$ $\sim [r_i A_1 ; \quad 0, A ; \ldots ; 0, A_{n-1} ; \quad 1 - r_i, A_n]$

Să notăm în această axiomă a continuității $[r_i, A_1 ; 0, A_2 ; \ldots, 0, A_{n-1} ; 1 - r_i, A_n] = \hat{A}_i$ și ca urmare avem $A_i \sim \hat{A}_i$.

(L—R.4) În experimentul $E = [p_1, A_1 ; \ldots ; p_i, A_i ; \ldots ; p_n, A_n]$, evenimentul A_i poate fi înlocuit cu alternativa $\hat{A}_i$, adică

$$[p_1, A_1 ; \ldots ; p_i, A_i ; \ldots ; p_n ; A_n] \sim$$

$$\sim [p_1, A_1 ; \ldots ; p_i, \hat{A}_i ; \ldots ; p_n, A_n]$$

Această axiomă, a echivalenței, arată că prin substituirea lui A_i cu $\hat{A}_i$ se obțin două experimente echivalente.

(L—R.5) Preferința și echivalența sînt tranzitive, adică

$$[p_1, A_1 ; \ldots ; p_i, A_i ; \ldots ; p_n, A_n] \sim$$

$$\sim [p_1, \hat{A}_1 ; \ldots ; p_i, \hat{A}_i ; \ldots ; p_n, \hat{A}_n]$$

$$[p_1, A_1 ; \ldots ; p_i, A_i ; \ldots ; p_n, \hat{A}_n] \sim$$

$$\sim [p_1, A_1 ; (1 - p), A_n]$$

unde $p = p_1 r_1 + p_2 r_2 + \ldots + p_n r_n$.
Această axiomă, a tranzitivității, extinde tranzitivitatea din prima axiomă.

(L—R.6) Experimentul $[r, A_1 ; \quad 1 - r, A_n]$ este preferat sau echivalent experimentului $[r', A_1 ; 1 - r', A_n]$ dacă și numai dacă $r \geqslant r'$

În această axiomă, a monotoniei, se introduce o preferință proporțională cu probabilitatea de realizare a primului rezultat al alternativei.

Definiția 4.1.2. Se numește utilitate în sensul axiomaticii Luce-Raiffa o funcție u definită pe mulțimea experimentelor $\{E | E = [p_1, A_1 ; \ldots ; p_n, A_n]\}$ cu proprietatea că dacă $E \succcurlyeq E'$ atunci $u(E) \geqslant u(E')$.

Propoziția 4.1.2. Dacă relația de preferință sau echivalență $\succcurlyeq$ satisface axiomele (L—R.1)— (L—R.6) atunci există numerele u_i asociate evenimentelor A_i astfel încît pentru două experimente $E = [p_1, A_1 ; \ldots ; p_n, A_n]$ și

$E' = [p'_1, A_1; \ldots; p'_n, A_n]$ relaţia dintre valorile $\sum_{k=1}^{n} p_k u_k$ şi $\sum_{k=1}^{n} p'_k u_k$ decide preferinţa experimentului.

Demonstraţie. Ca urmare a primelor cinci axiome E şi E' pot fi aduse la forma din axioma a şasea, conform căreia putem alege între ele, calculînd $\sum_{k=1}^{n} p_k u_k$ şi $\sum_{k=1}^{n} p'_k u_k$. Dacă prima sumă este mai mare preferăm E, iar în caz contrar preferăm E'. În caz de egalitate ele sînt echivalente.

Se spune că funcţia de utilitate este liniară dacă $u([p, E; 1-p, E']) = pu(E) + (1-p)u(E')$ pentru orice probabilitate $p \in (0,1)$ şi orice două experimente E şi E'.

Avem mai sus două axiomatici asupra utilităţii unui rezultat al unui experiment, sau asupra utilităţii unui experiment, sau altfel spus asupra utilităţii unui eveniment sau a unui cîmp de evenimente. Se poate observa că unele axiome apar ca naturale, altele mai puţin naturale şi rezultatul este oarecum explicabil prin însăşi natura conceptului de utilitate, căci preferinţele variază de la individ la individ şi chiar pentru un individ oarecare, de la un moment la altul. Apare chiar ca o regulă de prudenţă căutarea unui număr cît mai mic de ipoteze asupra utilităţii, general valabile şi suficient de elastice în formulare. Este într-un fel explicabilă dificultatea definirii unei măsuri a calităţii informaţiei, fie şi numai sub aspectul utilităţii acesteia, căci este greu şi adeseori chiar imposibil să se exprime numeric în mod unic ceea ce apare calitativ într-un experiment. Din acest motiv şi rezultatele privind măsurarea utilităţii informaţiei sînt extrem de puţine în raport cu măsurarea cantităţii de informaţie.

4.2. Entropia ponderată a lui S. Guiaşu (1968)

Pornind de la măsura cantităţii de informaţie introdusă de Shannon (1948) şi ţinînd cont de conceptul de utilitate axiomatizat de von Neumann şi Morgenstern (1944), Luce şi Raiffa (1957) şi alţii, S. Guiaşu a introdus în 1968 şi apoi a extins în 1971 conceptul de entropie ponderată, putîndu-se astfel diferenţia între ele calitativ două informaţii egale sau nu cantitativ. În linii mari este vorba de a ataşa fiecărei informaţii individuale o anumită pondere în totalul informaţiei furnizate de un experiment oarecare, pondere care poate

Scanned with OKEN Scanner

însemna importanța informației individuale sau utilitatea acesteia sau a evenimentului corespunzător. Dacă utilitatea în sensul larg von Neumann-Morgenstern sau Luce-Raiffa poate fi și negativă, în cazul teoriei lui S. Guiașu aceasta este totdeauna pozitivă și direct proporțională cu importanța sau semnificația cantității de informație corespunzătoare unui eveniment sau unui experiment.

Definiția 4.2.1. Fiind dat un experiment aleator finit urmînd repartiția $p_k \geqslant 0$, $\sum_{k=1}^{n} p_k = 1$, precum și numerele $u_k \geqslant 0$ direct proporționale cu importanța evenimentului de probabilitate p_k, se numește entropie ponderată expresia

$$\Gamma_n[u, p] = \Gamma_n[u_1, \ldots, u_n; p_1, \ldots, p_n] = -\sum_{k=1}^{n} u_k p_k \log p_k. \tag{4.2.1}$$

În modul în care entropia Shannon este definită, ea apare tot ca o informație totală ponderată. Entropia ponderată mai include în afara ponderii p_k ce reprezintă probabilitatea de obținere a cantității de informație $(-\log p_k)$ și ponderea u_k desemnînd importanța sau utilitatea unei unități din informația $(-\log p_k)$.

Propoziția 4.2.1. Entropia ponderată $\Gamma_n[u_1, \ldots, u_n; p_1, \ldots, p_n]$ satisface următoarele:

(P.1) este definită pentru orice repartiție p_k, $\sum_{k=1}^{n} p_k = 1$

(P.2) este simetrică în raport cu parametrii u_k și p_k, $1 \leqslant k \leqslant n$

(P.3) $\Gamma_n[u_1, \ldots, u_n; p_1 \ldots, p_n] \geqslant 0$

(P.4) dacă $p_{k_0} = 1$ și $p_k = 0$, $1 \leqslant k \leqslant n$, $k \neq k_0$, atunci $\Gamma_n[u_1, \ldots, u_n; p_1, \ldots, p_n] = 0$ oricare ar fi $\{u_k\}_{k=1}^{n}$

(P.5) dacă $p_i = 0$, $u_i \neq 0$ $(\forall)$ $i \in I$ și $p_j \neq 0$, $u_j = 0$ $(\forall)$ $j \in J$, $I \cap J = \emptyset$, $I \cup J = \{1, 2, \ldots, n\}$, atunci $\Gamma_n[u_1, \ldots, u_n; p_1, \ldots, p_n] = 0$

(P.6) dacă $u_1 = u_2 = \ldots = u_n = u$, atunci $\Gamma_n[u_1, \ldots, u_n; p_1, \ldots, p_n] = uH[p_1, \ldots, p_n]$

(P.7) $\Gamma_{n+1}[u_1, \ldots, u_n, u_{n+1}; p_1, \ldots, p_n, 0] = \Gamma_n[u_1, \ldots, u_n; p_1, \ldots, p_n]$

pentru orice $\{u_k\}_{k=1}^{n+1}$ și orice $p_k \geqslant 0$, $\sum_{k=1}^{n} p_k = 1$

(P.8) pentru orice $\lambda \geqslant 0$ avem

$$\Gamma_n[\lambda u_1, \ldots, \lambda u_n; p_1, \ldots, p_n] =$$

$$= \lambda \Gamma_n[u_1, \ldots, u_n; p_1, \ldots, p_n].$$

Demonstrație. Afirmațiile sînt imediate, pornind de la definiția 4.2.1.

Propoziția 4.2.2. Avem

$$\Gamma_{n+1}[u_1, \ldots, u_{n-1}, v_1, v_2; p_1, \ldots, p_{n-1}, q_1, q_2] =$$

$$= \Gamma_n[u_1, \ldots, u_n; p_1, \ldots, p_n] + p_n \Gamma_2\left[v_1, v_2; \frac{q_1}{p_n}, \frac{q_2}{p_n}\right] \tag{4.2.2}$$

unde

$$u_n = \frac{q_1 v_1 + q_2 v_2}{q_1 + q_2}; \quad q_1 + q_2 = p_n \tag{4.2.3}$$

Demonstrație. Pornind de la definiția 4.2.1 și notînd u_n și p_n ca în (4.2.3) avem

$$\Gamma_{n+1}[u_1, \ldots, u_{n-1}, v_1, v_2; p_1, \ldots, p_{n-1}, q_1, q_2] =$$

$$= -\sum_{k=1}^{n-1} u_k p_k \log p_k - v_1 q_1 \log q_1 - v_2 q_2 \log q_2 =$$

$$= -\sum_{k=1}^{n-1} u_k p_k \log p_k - u_n p_n \log p_n + u_n p_n \log p_n -$$

$$- v_1 p_1 \log p_1 - v_2 p_2 \log p_2 = \Gamma_n[u_1, \ldots, u_n; p_1, \ldots, p_n] +$$

$$+ (q_1 v_1 + q_2 v_2) \log p_n - v_1 q_1 \log q_1 - v_2 q_2 \log q_2 =$$

$$= \Gamma_n[u_1, \ldots, u_n; p_1, \ldots, p_n] +$$

$$+ p_n\left[-v_1 \frac{q_1}{p_n} \log \frac{q_1}{p_n} - v_2 \frac{q_2}{p_n} \log \frac{q_2}{p_n}\right] =$$

$$= \Gamma_n[u_1, \ldots, u_n; p_1, \ldots, p_n] + p_n \Gamma_2\left[v_1, v_2; \frac{q_1}{p_n}, \frac{q_2}{p_n}\right].$$

Scanned with OKEN Scanner

Deducem imediat

Propoziția 4.2.3. Dacă pentru orice $1 \leqslant k \leqslant n$,

a) $u_k = -\dfrac{p_k^{\alpha-1}}{\log p_k}$;

b) $u_k = -\dfrac{1}{(1-2^{1-\alpha})\log p_k}\left[1 - \sum_{k=1}^{n} p_k^{\alpha}\right]$;

c) $u_k = -\dfrac{1}{(1-\alpha)\log p_k} \log \sum_{k=1}^{n} p_k^{\alpha}, \quad \alpha > 0, \alpha \neq 1$ (4.2.4)

atunci

a) $\Gamma_n[u, p] = E_n^{\alpha}[p]$; b) $\Gamma_n[u, p] = I_n^{\alpha}[p]$;

c) $\Gamma_n[u, p] = H_n^{\alpha}[p]$ (4.2.5)

unde, $E_n^{\alpha}[p] = E_n^{\alpha}[p_1, \ldots p_n]$ este energia informațională Onicescu generalizată de ordin α dată de relația (3.5.55), $I_n^{\alpha}[p] = I_n^{\alpha}[p_1, \ldots, p_n]$ este alfa-entropia Behara-Nath dată de relația (3.5.50), iar $H_n^{\alpha}[p] = H_n^{\alpha}[p_1, \ldots, p_n]$ este entropia Rényi de ordin α dată de relația (3.5.16) pentru o repartiție completă.

Propoziția 4.2.4. Entropia ponderată este maximă dacă și numai dacă

$$p_k = 2^{\frac{\alpha}{u_k}} \cdot \frac{1}{e}, \quad 1 \leqslant k \leqslant n \tag{4.2.6}$$

unde α este soluția ecuației

$$\sum_{k=1}^{n} 2^{\frac{\alpha}{u_k}} = e \tag{4.2.7}$$

caz în care

$$[\max_{p}]\, \Gamma_n[u, p] = -\frac{\alpha}{e} + \frac{\log e}{e} \sum_{k=1}^{n} u_k \cdot 2^{\frac{\alpha}{u_k}} \tag{4.2.8}$$

Demonstrație. Aplicînd metoda multiplicatorilor lui Lagrange deducem imediat rezultatele menționate. Pentru comoditate în scriere putem lua $\hat{u}_k = u_k \log_2 e$.

Să remarcăm că pentru $u_k = u$, $1 \leqslant k \leqslant n$, obținem repartiția uniformă care maximizează entropia Shannon.

Pornind de la axiomaticile entropiei Shannon, S. Guiașu (1977) a dat o axiomatică a entropiei ponderate și ca urmare

a demonstrat unicitatea acesteia, prin analogie cu modul în care Fadeev a demonstrat unicitatea entropiei Shannon. Vom prezenta acest rezultat fără demonstrația corespunzătoare.

Propoziția 4.2.5. Fie șirul de funcții reale nenegative $\{\Gamma_n[u_1, \ldots u_n; p_1, \ldots, p_n]\}_{n=1}^{\infty}$ definite pe mulțimea $u_k \geqslant 0$, $p_k \geqslant 0$, $1 \leqslant k \leqslant n$, $\sum_{k=1}^{n} p_k = 1$.

Dacă

(G.1) $\Gamma_2[u_1, u_2; p, 1-p]$ este continuă în raport cu $p \in [0, 1]$

(G.2) $\Gamma_n[u_1, \ldots, u_n; p_1, \ldots, p_n]$ este simetrică în raport cu toate perechile (u_k, p_k), $1 \leqslant k \leqslant n$

(G.3) $\Gamma_{n+1}[u_1, \ldots, u_{n-1}, v_1, v_2; p_1, \ldots, p_{n-1}, q_1, q_2] =$

$$= \Gamma_n[u_1, \ldots, u_n; p_1, \ldots p_n] + p_n \Gamma_2\left[v_1, v_2; \frac{q_1}{p_n}, \frac{q_2}{p_n}\right]$$

cînd

$$u_n = \frac{q_1 v_1 + q_2 v_2}{q_1 + q_2}, \quad p_n = q_1 + q_2$$

(G.4) $\Gamma_n\left[u_1, \ldots, u_n; \frac{1}{n}, \ldots, \frac{1}{n}\right] = L(n) \cdot \frac{1}{n} \sum_{i=1}^{n} u_i,$

$L(n)$ fiind un număr pozitiv pentru orice $n > 1$

atunci

$$\Gamma_n[u_1, \ldots, u_n; p_1, \ldots, p_n] = -\lambda \sum_{k=1}^{n} u_k p_k \log p_k \quad (4.2.9)$$

unde λ este o constantă pozitivă arbitrară.

Definiția 4.2.2. Fiind dată o variabilă aleatoare X care ia un număr finit de valori nenegative x_k, $1 \leqslant k \leqslant n$, cu probabilitățile $p_k \geqslant 0$, $\sum_{k=1}^{n} p_k = 1$, atunci numim entropie ponderată a lui X expresia

$$\Gamma[X] = -\sum_{k=1}^{n} x_k p_k \log_2 p_k \quad (4.2.10)$$

Ca urmare a relației (4.2.10) deducem imediat

Scanned with OKEN Scanner

Propoziția 4.2.6. Pentru $\Gamma[X]$ dată de (4.2.10) sînt adevărate următoarele

1) dacă $X = k$, atunci $\Gamma[X] = 0$

2) dacă $Y = aX + b$, atunci $\Gamma[Y] = a\Gamma[X]$

3) dacă X și Y sînt independente atunci $\Gamma[X, Y] = M(X)\Gamma[Y] + M(Y)\Gamma[X]$ unde $M(X)$, $M(Y)$ reprezintă valorile medii ale variabilelor X și Y.

Considerînd acum numerele u_k din definiția 4.2.1 ca fiind măsuri ale utilității unei unități din cantitatea de informație $(-\log p_k)$, atunci utilitatea întregii cantități de informație furnizată de evenimentul $A_k \in \mathfrak{K}$ este $(-u_k \log p_k)$ și în consecință utilitatea medie totală a cantității de informație furnizată de un experiment cu n rezultate, sau a unui cîmp cu n evenimente elementare $\{A_k\}_{k=1}^n$, $A_k \cap A_j = \emptyset$, $k \neq j$, $\bigcup_{k=1}^{n} A_k = \Omega$, este

$$U(\Omega) = \sum_{k=1}^{n} p_k u(A_k) = \sum_{k=1}^{n} p_k(-u_k \log p_k) = -\sum_{k=1}^{n} u_k p_k \log p_k \tag{4.2.11}$$

adică tocmai entropia ponderată. Prin urmare putem spune că entropia ponderată poate fi socotită și ca măsură a utilității cantității de informație pe care o conține un cîmp de probabilitate finit, deci un experiment aleator oarecare cu un număr finit de rezultate posibile. În consecință toate proprietățile entropiei ponderate pot fi formulate în terminologia de utilitate, observînd totodată că noțiunea de entropie ponderată este mai generală decît cea de utilitate a informației și totodată decît chiar entropia Shannon. În sens larg, entropia ponderată poate fi considerată ca o măsură a calității informației furnizate de un anumit experiment, înțelegînd prin calitate a informației importanța, semnificația sau utilitatea acesteia. Scriind relația de preferință $E_1 \succcurlyeq E_2$, putem spune că informațional experimentul E_1 este preferabil lui E_2, sau că informațional E_1 este mai important decît E_2, sau că informația furnizată de E_1 este mai importantă sau mai utilă decît informația furnizată de E_2.

Din punct de vedere economic, ponderile u_k ar putea fi interpretate și drept cheltuieli de obținere a informației $(-\log p_k)$ ceea ce ar însemna că entropia ponderată ar avea

Scanned with OKEN Scanner

acum semnificația de cheltuială totală medie de obținere a tuturor informațiilor $(-\log p_k)_{k=1}^n$. În mod analog am putea spune că u_k reprezintă beneficiul realizat prin utilizarea într-un anume domeniu a informației $(-\log p_k)$ ceea ce ar însemna că $\Gamma_n[u_1, \ldots, u_n; p_1, \ldots, p_n]$ reprezintă beneficiul total mediu realizat de pe urma valorificării tuturor informațiilor $(-\log p_k)_{k=1}^n$, pentru care, propoziția 4.2.4 apare ca o problemă de extremum ce poate fi extinsă dacă apar și restricții matematico-economice suplimentare.

Scanned with OKEN Scanner

II. MĂSURA CORELAȚIEI

5. Problema măsurării corelației

5.1. Considerații generale

O preocupare importantă a statisticii matematice a constituit-o măsurarea gradului de dependență dintre două sau mai multe caracteristici aleatoare ale unei colectivități statistice. Primele rezultate în această direcție datează de la sfîrșitul secolului trecut și sînt legate de numele fondatorilor școlii anglo-saxone de statistică matematică sir Francis Galton (1822—1911) și Karl Pearson (1857—1936). În anii 1886—1890, Galton, antropolog de renume, preocupat de studiul fenomenelor de ereditate, avansează ideea „variației legate", idee pe care Pearson o va transpune în limbaj matematic adecvat, punînd bazele teoriei corelației, care, la rîndul ei, va fi riguros analizată și dezvoltată de către sir Ronald Aylmer Fisher (1890—1962), creatorul celei mai mari părți a statisticii matematice.

În ciuda unor încercări ce vizau unele remedii ale ansamblului de măsuri ale dependenței introduse de Galton, Pearson și Fisher, rezultatele acestora au rămas valabile și astăzi în practica statistică și constituie ceea ce numim pe scurt corelația clasică sau corelația Galton-Pearson.

Apariția teoriei matematice a informației în 1948 a deschis o nouă cale de studiere a dependenței și a condus la definirea măsurilor informaționale ale dependenței, de care însă poate fi legat și numele lui Fisher prin preocupările sale privind măsura informației, înainte de 1948 și după aceea. Pe linia definirii și studierii unor măsuri informaționale ale dependenței, pot fi remarcate mai multe lucrări între care cele ale spaniolilor M. Castans Camargo și Medina e Isabel (1955, 1956) menționate și utilizate de E. H. Linfoot (1957), sau cele elaborate de S. Kullback (1959), A. Rényi (1959), C. B. Bell

Scanned with OKEN Scanner

(1962), C. Rajski (1963), S.D. Silvey (1964), O. Onicescu (1966), S. Watanabe (1969), J. Aczel (1972), Y. Horibe (1973), S. Guiaşu (1973), I. Purcaru (1976, 1979, 1981 etc,) şi alţii.

5.2. Postulatele lui A. Rényi (1959)

Din ansamblul general de preocupări pentru definirea unor măsuri ale dependenţei, al căror scop final, aşa cum remarcă O. Onicescu, rămîne acelaşi : „identificarea unor noi forme de legi de dependenţă specific statistice şi admirabil adaptate la formele complexe şi variate pe care le oferă viaţa în diferitele sale manifestări", se poate desprinde un rezultat remarcabil şi anume încercarea de axiomatizare a conceptului de „bună măsură a dependenţei" pe care a dat-o A. Rényi în 1959. Cele şapte postulate ale lui Rényi au drept scop definirea unei măsuri etalon a dependenţei, ale cărei proprietăţi să apară cît mai naturale. Modelul axiomatizat de Rényi permite introducerea unei relaţii de preferinţă pe mulţimea indicatorilor dependenţei definiţi în decursul anilor şi în cele din urmă o clasificare a acestora prin prisma ideii de „bună măsură a dependenţei".

Fie $\{\Omega, \mathcal{K}, P\}$ un cîmp de probabilitate şi $\mathfrak{X}$ mulţimea tuturor variabilelor aleatoare definite pe acesta.

Definiţia 5.1. O funcţie $\delta : \mathfrak{X} \times \mathfrak{X} \to \mathbf{R}$ se numeşte „bună măsură a dependenţei dintre două variabile X_1 şi $X_2 \in \mathfrak{X}$, în sensul lui Rényi, dacă satisface următoarele axiome :

(R.I) este definită pentru orice X_1 şi $X_2 \in \mathfrak{X}$, neconstante cu probabilitatea unu

(R.II) este simetrică în raport cu X_1 şi X_2, adică

$$\delta(X_1, X_2) = \delta(X_2, X_1) \tag{5.2.1}$$

(R.III) $0 \leqslant \delta(X_1, X_2) \leqslant 1$ pentru orice (X_1, X_2)

(R.IV) $\delta(X_1, X_2) = 0$ dacă şi numai dacă X_1 şi X_2 sînt independente

(R.V) $\delta(X_1, X_2) = 1$ dacă există o dependenţă funcţională între X_1 şi X_2, adică $X_2 = \varphi_1(X_1)$ sau $X_1 = \varphi_2(X_2)$, φ_1 şi φ_2 fiind funcţii reale biunivoc definite pe axa reală

(R.VI) este invariantă faţă de orice transformare biunivocă

Scanned with OKEN Scanner

a lui X_1 și X_2, adică

$$\delta(\varphi_1(X_1),\ \varphi_2(X_2)) = \delta(X_1, X_2) \tag{5.2.2}$$

(R.VII) oricare ar fi $\varphi_1, \varphi_2 : \mathbf{R} \to \mathbf{R}$, definite în mod biunivoc dacă vectorul (X_1, X_2) este repartizat normal atunci

$$\delta(X_1, X_2) = |\rho(X_1, X_2)| \tag{5.2.3}$$

unde $\rho(X_1, X_2)$ este coeficientul de corelație clasic.

Să remarcăm că postulatul al cincilea nu impune și condiția „numai dacă", Rényi considerînd că proprietatea ar fi prea restrictivă. Ca urmare, dacă X_1 și X_2 sînt funcțional dependente atunci $\delta(X_1, X_2) = 1$, dar dacă $\delta(X_1, X_2) = 1$ atunci nu înseamnă că X_1 și X_2 sînt neapărat funcțional dependente. Putem observa de asemenea că postulatele lui Rényi sînt date pentru măsura dependenței dintre două variabile. Deoarece în practica statistică apar probleme privind dependența a mai mult de două variabile, este necesară o extensiune adecvată a axiomaticii lui Rényi.

Analizînd pe rînd măsurile dependenței probabilistice ce vor fi prezentate în cele ce urmează, fără a ne propune o analiză critică a acestora, vom putea remarca, din proprietățile pe care acestea le au, care anume poate fi socotită ca „bună măsură" a dependenței în sensul lui Rényi. S-au făcut și unele încercări de modificare a unora dintre postulate, în speță cel de al șaptelea și ca urmare clasa măsurilor de dependență ce verifică noile postulate (în sensul lui Rényi-Bell) poate fi mărită considerabil.

6. Măsuri clasice

6.1. Corelația Galton-Pearson (1901)

Numele de corelație Galton-Pearson poate fi atribuit aproape firesc corelației clasice pentru că principalele rezultate de început de aici aparțin acestor personalități de marcă în lumea științei, chiar dacă ulterior alți matematicieni sau oameni de știință din alte domenii au adus însemnate contribuții la dezvoltarea teoriei corelației. Ideile generale fuseseră formulate. Localizarea în timp pentru anul 1901

este făcută pentru a ilustra că publicarea unor rezultate importante privind corelaţia şi regresia este strîns legată de data apariţiei primului număr al revistei „Biometrika", fondată de Galton şi Pearson în principal, chiar dacă anumite idei şi rezultate fuseseră făcute cunoscute înainte de 1901 sau au fost publicate după 1901, acestea din urmă îmbunătăţind de fapt ceea ce luase naştere în acel prim an al secolului al XX-lea. Întîmplător sau nu, la început de veac se năştea biometria, la frontiera dintre matematică, statistică şi antropologie, iar „Biometrika" era actul său de naştere!

6.1.1. Coeficient de corelaţie simplă

În acest caz este vorba despre corelaţia liniară dintre două componente ale unui vector aleator. Fie $X = (X_1, X_2)$ un vector aleator discret (sau continuu), cu densitatea de repartiţie

$$p(x_1, x_2) \geqslant 0, \sum_{x_1} \sum_{x_2} p(x_1, x_2) = 1 \left(\text{sau} \int_{\mathbf{R}^2} p(x_1, x_2)\, dx_1\, dx_2 = 1 \right).$$

Definiţia 6.1.1. Se numeşte moment de ordinul (r_1, r_2) al vectorului aleator (X_1, X_2) valoarea medie a variabilei $X_1^{r_1}, X_2^{r_2}$, dată de relaţia

$$M_{r_1 r_2} = M[X_1^{r_1} \cdot X_2^{r_2}] = \sum_{x_1} \sum_{x_2} x_1^{r_1} \cdot x_2^{r_2}\, p(x_1, x_2)$$

$$\left(\text{sau} \int_{\mathbf{R}^2} x_1^{r_1} \cdot x_2^{r_2}\, p(x_1, x_2)\, dx_1\, dx_2 \right) \qquad (6.1.1)$$

Dacă $r_1 = r_2 = 1$ obţinem valoarea medie a variabilei produs $X_1 X_2$, iar dacă $r_1 = 1$, $r_2 = 0$ sau $r_1 = 0$, $r_2 = 1$, valoarea medie a lui X_1 sau a lui X_2.

Definiţia 6.1.2. Se numeşte moment centrat de ordinul (r_1, r_2) al vectorului (X_1, X_2), valoarea medie a variabilei $(X_1 - M_{10})^{r_1} \cdot (X_1 - M_{01})^{r_2}$ adică

$$m_{r_1 r_2} = \sum_{x_1} \sum_{x_2} (x_1 - M_{10})^{r_1} (x_2 - M_{01})^{r_2}\, p(x_1, x_2)$$

$$\left(\text{sau} \int_{\mathbf{R}^2} (x_1 - M_{10})^{r_1} \cdot (x_2 - M_{01})^{r_2}\, p(x_1, x_2)\, dx_1\, dx_2 \right). \qquad (6.1.2)$$

Scanned with OKEN Scanner

Dacă $r_1 = 2$, $r_2 = 0$ sau $r_1 = 0$, $r_2 = 2$ obținem $m_{20} = \sigma_1^2$ sau $m_{02} = \sigma_2^2$, iar dacă $r_1 = r_2 = 1$, obținem $m_{11} = M[(X_1 - M_{10})(X_2 - M_{01})]$.

Definiția 6.1.3. Se numește covarianță a variabilelor X_1 și X_2 expresia

$$\operatorname{cov}(X_1, X_2) = M[(X_1 - M_{10})(X_2 - M_{02})] \qquad (6.1.3)$$

și coeficient de corelație Galton-Pearson al acestora expresia

$$\rho(X_1, X_2) = \frac{\operatorname{cov}(X_1, X_2)}{\sigma_1 \cdot \sigma_2}. \qquad (6.1.4)$$

Dacă în locul vectorului (X_1, X_2) vom considera vectorul de componente standardizate $\left(\frac{X_1 - M_{10}}{\sigma_1}, \frac{X_2 - M_{01}}{\sigma_2}\right)$ atunci relația (6.1.4) poate fi scrisă sub forma

$$\rho(X_1, X_2) = \operatorname{cov}\left[\frac{X_1 - M_{10}}{\sigma_1}, \frac{X_2 - M_{01}}{\sigma_2}\right]. \qquad (6.1.5)$$

Propoziția 6.1.1. Pentru coeficientul de corelație $\rho(X_1, X_2)$ sînt adevărate următoarele:

1) este definit pentru orice vector aleator (X_1, X_2), neconstant cu probabilitatea unu
2) $\rho(X_1, X_2) = \rho(X_2, X_1)$ pentru orice (X_1, X_2)
3) $-1 \leqslant \rho(X_1, X_2) \leqslant 1$ pentru orice (X_1, X_2)
4) dacă X_1 și X_2 sînt independente atunci $\rho(X_1, X_2) = 0$ nu și reciproc
5) dacă $\rho^2(X_1, X_2) = 1$, atunci $X_2 = a_1 X_1 + b_1$ sau $X_1 = a_2 X_2 + b_2$
6) dacă $X_1 = a_2 X_2 + b_2$ sau $X_2 = a_1 X_1 + b_1$, atunci $\rho^2(X_1, X_2) = 1$
7) dacă $Y_1 = a_1 X_1 + b_1$ și $Y_2 = a_2 X_2 + b_2$, $a_1 > 0$, $a_2 > 0$, atunci $\rho(Y_1, Y_2) = \rho(X_1, X_2)$

Demonstrație. Proprietățile 1) și 2) sînt evidente.
Fie forma pătratică pozitiv definită

$$\int_{R^2} [\lambda(x_1 - M(X_1)) + \mu(x_2 - M(X_2))]^2 p(x_1, x_2)\, dx_1\, dx_2 =$$

$$= \lambda^2 \sigma_1^2 + 2\lambda\mu\rho\sigma_1\sigma_2 + \mu^2\sigma_2^2 \geqslant 0$$

Scanned with OKEN Scanner

a cărei matrice are determinantul pozitiv, cu expresia

$$0 \leqslant \begin{vmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{vmatrix} = \sigma_1^2\,\sigma_2^2(1-\rho^2)$$

de unde rezultă că $1 - \rho^2 \geqslant 0$ şi deci proprietatea 3). Dacă $\rho^2 = 1$, atunci determinantul este zero şi ca urmare există două numere λ_0 şi μ_0 încît să avem relaţia $\lambda_0(X_1 - M(X_1)) + \mu_0(X_2 - M(X_2)) = 0$ ceea ce arată că între X_1 şi X_2 există o dependenţă liniară şi deci proprietatea 5). Proprietăţile 4), 6) şi 7) rezultă imediat prin calcul direct, iar în ceea ce priveşte reciproca proprietăţii 4) următorul exemplu este suficient. Fie $X_1 : \begin{pmatrix} -1 & 0 & 1 \\ 1/4 & 1/2 & 1/4 \end{pmatrix}$ şi $X_2 = X_1^2$ care exprimă o dependenţă funcţională între X_1 şi X_2. Determinăm cu uşurinţă repartiţia comună a vectorului (X_1, X_2) şi obţinem $\rho(X_1, X_2) = 0$, cu toate că X_1 şi X_2 nu sînt independente.

Să constatăm că $\rho(X_1, X_2)$ furnizează destule informaţii asupra dependenţei liniare dintre X_1 şi X_2, dar nu complete căci spre exemplu, dacă $\rho(X_1, X_2) = 0$, nu putem afirma nimic despre X_1 şi X_2, ele putînd fi dependente sau independente. În acest caz spunem că X_1 şi X_2 sînt în general necorelate. Observăm de asemenea că el nu satisface postulatele lui Rényi.

Fie acum $X = (X_1, \ldots, X_n)$ un vector aleator n-dimensional, discret sau continuu şi să notăm cu $M_i = M(X_i)$ şi $\sigma_i^2 = D^2(X_i)$, valoarea medie şi respectiv dispersia componentei X_i, $1 \leqslant i \leqslant n$, iar cu $\sigma_{ij} = \operatorname{cov}(X_i, X_j)$, covarianta perechii (X_i, X_j), $1 \leqslant i, j \leqslant n$.

Evident că $\sigma_{ii} = \sigma_i^2$ şi $\rho(X_i, X_j) = \rho_{ij} = \dfrac{\sigma_{ij}}{\sigma_i \cdot \sigma_j}$.

Definiţia 6.1.4. Matricea $\Sigma(X) = (\sigma_{ij})$ se numeşte matrice covariantă sau matrice de covarianţă corespunzătoare lui X, iar matricea $\mathcal{R}(X) = (\rho_{ij})$ se numeşte matrice de corelaţie a lui X.

Ca urmare a propoziţiei 6.1.1 deducem

Propoziţia 6.1.2. Pentru matricea $\Sigma(X)$ sînt adevărate afirmaţiile

1) este definită pentru orice vector aleator

2) este simetrică, adică $\sigma_{ij} = \sigma_{ji}$, pentru orice $1 \leqslant i, j \leqslant n$

3) $\sigma_{ij}^2 \leqslant \sigma_i^2 \cdot \sigma_j^2$ pentru orice $1 \leqslant i, j \leqslant n$

Scanned with OKEN Scanner

4) dacă $\det \Sigma(X) = |\Sigma(X)| = 0$ atunci între componentele lui X există cel puțin o relație liniară

5) dacă X are componentele independente două cîte două atunci

$$\sigma_{ij} = 0 \text{ pentru } i \neq j \text{ și } |\Sigma(X)| = \prod_{i=1}^{n} \sigma_i^2$$

Ca o consecință a propoziției 6.1.2 avem

Propoziția 6.1.3. Pentru matricea $\mathcal{R}(X)$ sînt adevărate următoarele

1) este definită pentru orice X pentru care $\sigma_i > 0$ pentru orice $1 \leqslant i \leqslant n$

2) este simetrică, adică $\rho_{ij} = \rho_{ji}$

3) $-U \leqslant \mathcal{R}(X) \leqslant U$, U fiind matricea cu toate elementele egale cu unu, cu egalitate dacă pentru orice $1 \leqslant i, j \leqslant n$, $X_j = a_i X_i + b_i$

4) dacă $|\mathcal{R}(X)| = 0$, atunci între componentele lui X avem cel puțin o relație liniară

5) dacă X are componentele independente două cîte două, atunci
$\mathcal{R}(X) = E$, E fiind matricea unitate de ordinul n

6) $|\Sigma(X)| = |\mathcal{R}(X)| \prod_{i=1}^{n} \sigma_i^2$

6.1.2. Rapoarte de corelație

Să considerăm variabila aleatoare condiționată $(X_1/X_2) = \{X_1/x_2\}\, x_2$ a cărei densitate de repartiție este $p(x_1/x_2)$ pentru fiecare componentă X_1/x_2.

Definiția 6.1.5. Se numește regresie a lui X_1 asupra lui X_2 valoarea medie a variabilei condiționate X_1/x_2 dată de relația

$$M(X_1/x_2) = \int_{\mathbf{R}} x_1\, p(x_1/x_2)\, dx_1 \text{ (sau } \sum_{x_1} x_1\, p(x_1/_2)) \qquad (6.1.6)$$

după cum (X_1, X_2) sînt continue sau discrete.

Să observăm că $M(X_1/x_2) = f(x_2)$ și deci $M(X_1/X_2) = \{M(X_1/x_2)\}x_2$ adică $M(X_1/X_2)$ apare ca mulțimea tuturor valorilor medii condiționate $M(X_1/x_2)$. Mulțimea perechilor

Scanned with OKEN Scanner

$(x_2,\ M(X_1/x_2))$ definește o curbă numită curbă de regresie a lui X_1 asupra lui X_2 și după cum aceasta este o dreaptă, parabolă sau altă curbă, putem spune că avem o dreaptă, o parabolă sau o altă curbă de regresie.

Relația (6.1.6) poate fi extinsă considerînd variabila aleatoare condiționată $\varphi(X_1)/x_2$ și corespunzător $\varphi(X_1)/X_2$.

Să luăm $\varphi(X_1) = [X_1 - M(X_1/x_2)]^2$. Valoarea medie a lui $\varphi(X_1)$ se numește dispersie a variabilei aleatoare condiționată X_1/x_2. Deducem imediat că $M[M(X_1^k/X_2)] = M(X_1^k)$. Într-adevăr,

$$M[M(X_1^k/X_2)] = \int_R p(x_2) M(X_1^k/x_2)\mathrm{d}x_2 =$$

$$= \int_R p(x_2)\left[\int_R p(x_1/x_2)x_1^k\,\mathrm{d}x_1\right]\mathrm{d}x_2 =$$

$$= \iint_{R^2} x_1^k p(x_1,\ x_2)\mathrm{d}x_1\mathrm{d}x_2 = M(X_1^k).$$

Să notăm cu $\sigma^2(X_1/X_2)$ sau $D^2(X_1/x_2)$ această valoare medie, adică

$$D^2(X_1/x_2) = M[(X_1 - M(X_1/x_2))^2] \qquad (6.1.7)$$

Putem vorbi acum de variabila aleatoare ale cărei valori, pentru fiecare x_2, sînt date de relația (6.1.7) și pe care să o notăm cu $D^2(X_1/X_2)$. Să considerăm prin urmare variabila

$$\psi_2(X_2) = \frac{D^2(X_1/X_2)}{D^2(X_1)},\ D^2(X_1) > 0 \qquad (6.1.8)$$

Definiția 6.1.6. Se numește raport de corelație a lui X_1 asupra lui X_2 valoarea medie a variabilei $\psi_2(X_2)$ și anume

$$\eta^2_{1.2}(X_1, X_2) = \frac{M[D^2(X_1/X_2)]}{D^2(X_1)} \qquad (6.1.9)$$

Deducem fără dificultate că $D^2(X_1) \geqslant M[D^2(X_1/X_2)]$. Într-adevăr avem $D^2(X_1/X_2) = M(X_1^2/X_2) - M^2(X_1/X_2)$ și ca urmare

$$M[D^2(X_1/X_2)] = M[M(X_1^2/X_2)] - M[M^2(X_1/X_2)] =$$

$$= M(X_1^2) - M[M^2(X_1/X_2)] =$$

Scanned with OKEN Scanner

$$= D^2(X_1) + M^2(X_1) - M[M^2(X_1/X_2)] =$$

$$= D^2(X_1) - M\{[M(X_1/X_2) - M(X_1)]^2\} \leqslant D^2(X_1)$$

cu egalitate dacă și numai dacă $M\{[M(X_1/X_2) - M(X_1)]^2\} = 0$ ceea ce are loc dacă și numai dacă X_1 și X_2 sînt independente.

Remarcăm de asemenea că pentru orice X_1 și X_2 are loc inegalitatea $M^2(X_1) \leqslant M[M^2(X_1/X_2)]$. Totodată observăm că $M[D^2(X_1/X_2)] = 0$ implică $D^2(X_1/x_2) = 0$ pentru fiecare x_2, ceea ce înseamnă că $X_1 = \xi(X_2)$ și reciproc.

Ca urmare a ace tor rezultate putem formula

Propoziția 6.1.4. Raportul de corelație $\eta^2_{1.2}(X_1, X_2)$ satisface următoarele :

1) este definit pentru orice vector aleator (X_1, X_2) cu $D^2(X_1) > 0$
2) $\eta^2_{1.2}(X_1, X_2) \neq \eta^2_{2.1}(X_2, X_2)$
3) $0 \leqslant \eta^2_{1.2}(X_1, X_2) \leqslant 1$ pentru orice (X_1, X_2)
4) $\eta^2_{1.2}(X_1, X_2) = 0$ dacă și numai dacă $D^2(X_1/X_2) = 0$, adică $X_1 = \xi(X_2)$
5) $\eta^2_{1.2}(X_1, X_2) = 1$ dacă și numai dacă X_1 și X_2 sînt independente

Obișnuiți cu proprietățile coeficientului de corelație $\rho(X_1, X_2)$ și ținînd cont de postulatele lui Rényi, am putea considera drept raport de corelație și mărimea

$$\hat{\eta}^2_{1.2}(X_1, X_2) = 1 - \eta^2_{1;2} \qquad (6.1.10)$$

care schimbă între ele proprietățile 4) și 5) din propoziția 6.1.4 și fiind deci egal cu zero în cazul independenței și egal cu unu în cazul dependenței funcționale.

Tot sub numele de raport de corelație a lui X_1 asupra lui X_2 putem considera și expresia

$$\Theta^2_{1.2}(X_1, X_2) = \frac{D^2[M(X_1/X_2)]}{D^2(X_1)} \qquad (6.1.11)$$

Propoziția 6.1.5. Pentru raportul de corelație $\Theta^2_{1.2}(X_1, X_2)$ sînt adevărate afirmațiile :

1) este definit pentru orice (X_1, X_2) cu $D^2(X_1) > 0$
2) $\Theta^2_{1.2}(X_1, X_2) \neq \Theta^2_{2.1}(X_2, X_1)$

Scanned with OKEN Scanner

3) $0 \leqslant \Theta^2_{1.2}(X_1, X_2) \leqslant 1$

4) $\Theta^2_{1.2}(X_1, X_2) = 0$ dacă X_1 și X_2 sînt independente

5) $\Theta^2_{1.2}(X_1, X_2) = 1$ dacă și numai dacă $X_1 = \psi(X_2)$

6) facă $M(X_1)$ și $D^2(X_1)$ există, atunci pentru orice (X_1, X_2) avem

$$\Theta^2_{1.2}(X_1, X_2) = \sup_{\xi} |\rho[X_1, \xi(X_2)]| \qquad (6.1.12)$$

unde $\xi : \mathbf{R} \to \mathbf{R}$ este definită în mod biunivoc astfel că $M(\xi(X_2))$ și $D^2(\xi(X_2))$ există, avînd egalitate obișnuită dacă și numai dacă

$$\xi(X_2) = aM(X_1/X_2) + b, \qquad a \neq 0 \quad \text{și} \quad b \in \mathbf{R}.$$

7) dacă (X_1, X_2) este repartizat normal, atunci

$$\Theta^2_{1.2}(X_1, X_2) = \rho^2(X_1, X_2) \qquad (6.1.13)$$

8) $\eta^2_{1.2}(X_1, X_2) + \Theta^2_{1.2}(X_1, X_2) = 1$ pentru orice (X_1, X_2)

Demonstrație. Proprietățile 1) și 2) sînt evidente. Deoarece

$$D^2[M(X_1/X_2)] = M[M^2(X_1/X_2)] - M^2[M(X_1/X_2)] =$$

$$= M[M^2(X_1/X_2)] - M^2(X_1)$$

$$M[D^2(X_1/X_2)] = M[M(X_1^2/X_2) - M^2(X_1/X_2)] =$$

$$= M(X_1^2) - M[M^2(X_1/X_2)]$$

rezultă prin însumare că

$$D^2[M(X_1/X_2)] + M[D^2(X_1/X_2)] = D^2(X_1)$$

de unde deducem proprietățile 3) și 8). Să observăm că $D^2[M(X_1/X_2)] = 0$ dacă și numai dacă $M(X_1/X_2) =$ constant, ceea ce are loc dacă și numai dacă X_1 și X_2 sînt independente și ca urmare rezultă proprietatea 4). Să observăm că $D^2[M(X_1/X_2)] = D^2(X_1)$ implică $M[D^2(X_1/X_2)] = 0$ ceea ce este adevărat dacă și numai dacă $X_1 = \psi(X_2)$ de unde rezultă și proprietatea 5). Pentru demonstrația proprietății 6) să presupunem, fără a micșora generalitatea pro-

Scanned with OKEN Scanner

blemei că $M(X_1) = M[\xi(X_2)] = 0$ și că $D^2(X_1) = D^2[\xi(X_2)] = 1$. Totodată să observăm că $M(X_1\xi(X_2)) = \iint_{R^2} x_1\,\xi(x_2) \times$

$\times p(x_1, x_2)\,dx_1\,dx_2 = \int_R [\xi(x_2)p(x_2) \int_R x_1\,p(x_1/x_2)\,dx_1]\,dx_2 =$

$= M[M(X_1/X_2)\,\xi(X_2)$ și ca urmare, aplicînd și inegalitatea lui Schwarz, avem $\rho^2(X_1, \xi(X_2)) = M^2[M(X_1/X_2)\xi(X_2)] \leqslant$ $\leqslant M[M^2(X_1/X_2)]\,M[\xi^2(X_2))] = D^2[(M(X_1/X_2)]$ cu egalitate dacă și numai dacă $\xi(X_2) = aM(X_1/X_2) + b$, $a \neq 0$, $b \in \mathbf{R}$ și deci are loc și proprietatea 6). Dacă X_1 și X_2 sînt normal repartizate atunci deducem că $M(X_1/x_2) = \rho \frac{\sigma_1}{\sigma_2} x_2$ și ca urmare $D^2[M(X_1/X_2)] = \rho^2 \frac{\sigma_1^2}{\sigma_2^2} \sigma_2^2 = \rho^2 \sigma_1^2$ ceea ce implică $\Theta^2_{1,2} = \frac{\rho^2 \sigma_1^2}{\sigma_1^2} = \rho^2$, iar proprietatea 7) este verificată și odată cu ea și propoziția.

Ca urmare a proprietății 6) deducem că dacă $\xi(X_2) = aM(X_1/X_2) + b$ atunci, pentru cazul particular $M(X_1) = M[\xi(X_2)] = 0$ și $D^2(X_1) = D^2[\xi(X_2)] = 1$ obținem $b = 0$ și $\frac{1}{a} = D[M(X_1/X_2)]$ și în consecință

$$\xi(X_2) = \frac{M(X_1/X_2)}{D[M(X_1/X_2)]}.$$

Să observăm, ca urmare a proprietății 8) că relațiile (6.1.10) și (6.1.11) coincid. De asemenea, putem remarca că relația (6.1.9) poate fi ușor generalizată. Astel, considerînd variabila aleatoare condiționată $X_1/(x_2, \ldots, x_n)$ și variabila $\varphi(X_1) = [X_1 - M(X_1/(x_1, \ldots, x_n))]^2$, putem defini raportul de corelație a lui X_1 asupra vectorului $(X_2, \ldots, X_n)$ sau altfel spus al lui X_1 asupra lui $(X - X_1)$ ca fiind

$$\eta^2_{1.2\ldots n}[X_1(X - X_1)] = \frac{M[M(\varphi(X_1))]}{D^2(X_1)} = \frac{M[D(X_1/(X - X_1))]}{D^2(X_1)} \tag{6.1.14}$$

Ca urmare a relației (6.1.14), pentru un vector aleator n-dimensional X, se poate determina raportul de corelație al unei componente asupra oricărui grup de alte componente ale lui X.

Scanned with OKEN Scanner

În literatură, rapoartele de corelație mai sînt denumite și indici ai corelației neliniare sau curbilinii, spre deosebire de coeficientul de corelație liniară $\rho(X_1, X_2)$.

Să observăm că nici rapoartele de corelație nu satisfac postulatele lui Rényi.

6.1.3. Coeficient de corelație multiplă

Este vorba despre corelația dintre o componentă a lui X și un grup de alte componente ale acestuia ca și în relația (6.1.14), dar în ipoteza dependenței liniare dintre acestea.

Dacă generalizăm relația (6.1.6) luînd în locul variabilei X_2 vectorul $(X_2, X_3, \ldots, X_n)$ atunci putem da

Definiția 6.1.7. Valoarea medie a variabilei aleatoare condiționate $X_1/(x_2, \ldots, x_n)$ dată de relația

$$M[X_1/(x_2, \ldots, x_n)] = \sum_{x_1} x_1 p[x_1/(x_2, \ldots, x_n)]$$

$$\left(\text{sau} \int_R x_1 \, p[x_1/(x_2, \ldots, x_n)] \, dx_1\right) \qquad (6.1.15)$$

se numește regresia variabilei X asupra vectorului $(X_2, \ldots, X_n)$. Evident că $M[X_1/(x_2, \ldots, x_n)]$ este o funcție de $(n-1)$ variabile. Ea se numește funcție de regresie, iar graficul ei suprafață de regresie a lui X_1 față de $(X - X_1)$. Dacă această funcție este liniară, se spune că avem o regresie liniară. Să presupunem acest lucru, deci că

$$X_1 = \sum_{k=2}^{n} a_k X_k + a_1 \qquad (6.1.16)$$

relație care definește hiperplanul de regresie al lui X_1 asupra lui $(X - X_1)$ și care pentru $n = 2$ se reduce la dreapta de regresie (Mihoc (1966, 1970)). Aplicînd metoda celor mai mici pătrate pentru determinarea parametrilor a_k, $1 \leqslant k \leqslant n$, sîntem conduși la sistemul de ecuații liniare

$$a_1 = M(X_1) - \sum_{k=2}^{n} a_k M(X_k) \, ; \, A \cdot a = b \qquad (6.1.17)$$

unde $A = (\sigma_{ij})$, $2 \leqslant i, j \leqslant n$, $a' = (a_2, \ldots, a_n)$ și $b' = (\sigma_{1j})$, $2 \leqslant j \leqslant n$, pentru care, în ipoteza $|A| \neq 0$, deci cînd componentele lui X nu sînt dependente liniar două cîte două,

Scanned with OKEN Scanner

obținem soluția

$$\begin{cases} a_1 = M(X_1) - \sum_{k=2}^{n} a_k M_k \\ a_k = \dfrac{1}{|A|} \sum_{i=2}^{n} \sigma_{1i} A_{ik} \end{cases} \tag{6.1.18}$$

unde A_{ik} este complementul algebric al lui σ_{ik} din determinantul $|A|$ (după cum vectorul $b = (\sigma_{1i})$ substituie coloana $k = 2,\ldots,n$ din $|A|$ pentru obținerea soluției cu ajutorul determinanților). Ecuația hiperplanului de regresie (6.1.16) devine acum

$$x_1 - M_1 = \frac{1}{|A|} \sum_{j=2}^{n} \sum_{k=2}^{n} \sigma_{1i} A_{ik}(x_k - M_k) \tag{6.1.19}$$

și conține punctul $(M_1, M_2,\ldots,M_n)$ numit centru de greutate al repartiției vectorului aleator X (Mihoc (1966, 1970)).

Să considerăm acum variabila aleatoare

$$Z = \frac{1}{|A|} \sum_{i=2}^{n} \sum_{k=2}^{n} \sigma_{1i} A_{jk}(x_k - M_k). \tag{6.1.20}$$

Definiția 6.1.8. Se numește coeficient de corelație multiplă liniară a lui X_1 cu $(X_2,\ldots,X_n)$, sau a lui X_1 cu $(X - X_1)$, și se notează cu $\rho_{1.23\ldots n}$, coeficientul de corelație al variabilelor aleatoare X_1 și Z, adică

$$\rho_{1.23\ldots n} = \rho(X_1, Z) = \frac{\operatorname{cov}(X_1, Z)}{D(X_1)D(Z)}. \tag{6.1.21}$$

Deoarece (Mihoc (1966, 1970)

$$\operatorname{cov}(X_1, Z) = D^2(Z) = \frac{1}{|A|} \sum_{i=2}^{n} \sum_{k=2}^{n} \sigma_{1i}\, \sigma_{1k} A_{ik} \tag{6.1.22}$$

se poate deduce că

$$\rho_{1.23\ldots n}(X_1, (X - X_1)) = \rho_{1.23\ldots,n} = \sqrt{\frac{\sum_{j=2}^{n} \sum_{k=2}^{n} \sigma_{1i}\, \sigma_{1k} A_{ik}}{\sigma_{11}|A|}} \tag{6.1.23}$$

sau sub o altă formă (Mihoc (1966))

$$\rho^2_{1.23\ldots,n} = 1 - \frac{|\Sigma(X)|}{\sigma_{11}|A|} \tag{6.1.24}$$

Scanned with OKEN Scanner

Se observă uşor că pentru $n = 2$ relaţiile (6.1.4) şi (6.1.24) coincid. Din cele de mai sus rezultă

Propoziţia 6.1.6. Pentru coeficientul de corelaţie multiplă $\rho_{1.23\ldots n}$ sînt adevărate următoarele afirmaţii

1) este definit pentru orice vector aleator X cu $|A| \neq \neq 0$ şi $\sigma_{11} > 0$

2) nu este simetric în raport cu componentele lui X

3) $\rho^2_{1.23.,.n} \leqslant 1$

4) $\rho^2_{1.23\ldots n} = 0$ dacă şi numai dacă $|\Sigma(X)| = \sigma_{11}|A|$, adică, dacă şi numai dacă $X_2, X_3, \ldots, X_n$ sînt necorelate două cîte două

5) $\rho^2_{1.23\ldots n} = 1$ dacă şi numai dacă $|\Sigma(X)| = 0$, ceea ce se întîmplă dacă între oricare două componente ale lui X există o dependenţă liniară.

6.1.4. Coeficient de corelaţie parţială

Să considerăm iarăşi relaţia (6.1.16) în care a_k, $1 \leqslant \leqslant k \leqslant n$, sînt coeficienţii de regresie determinaţi de relaţia (6.1.18). Fiecare dintre ei indică existenţa unei corelaţii între X_1 şi variabila corespunzătoare coeficientului, cînd celelalte rămîn constante (Mihoc, (1966)).

Definiţia 6.1.9. Fiind dat vectorul aleator n-dimensional X, corelaţia dintre X_1 şi X_k, $2 \leqslant k \leqslant n$, cînd toate celelalte rămîn constante, se numeşte corelaţie parţială a variabilelor X_1 şi X_k.

Să considerăm acum variabilele aleatoare $Y_k = X_k - M_k$ şi $Y_1 = a_1 + \sum_{k=2}^{n} a_k Y_k$ şi să determinăm coeficienţii de regresie a_k, $1 \leqslant k \leqslant n$ pentru care $M[Y^2_{1.23\ldots n}] = D^2[Y^2_{1.23\ldots n}] = \sigma^2_{1.23\ldots n}$ este minimă, unde

$$Y_{1;23\ldots n} = Y_1 - \left[a_1 + \sum_{k=2}^{n} a_k Y_k \right]. \qquad (6.1.25)$$

Aplicînd metoda celor mai mici pătrate pentru determinarea coeficienţilor a_k pentru care $D^2[Y^2_{1.23\ldots n}]$ este minimă, sîntem conduşi la sistemul de ecuaţii

$$M[Y_k \cdot Y_{1;23.;.n}] = 0, \quad 2 \leqslant k \leqslant n \qquad (6.1.26)$$

din care obţinem pe a_k, $2 \leqslant k \leqslant n$.

Scanned with OKEN Scanner

Să notăm soluţia corespunzătoare lui a_k cu

$$\hat{a}_k = a_{1k.23\ldots(k-1)(k+1)\ldots n}, \quad 2 \leqslant k \leqslant n \tag{6.1.27}$$

ţinînd cont că aceasta este valoarea corelaţiei dintre X_1 şi X_k atunci cînd celelalte sînt constante. Ca urmare, avem

$$Y_1 = \sum_{k=2}^{n} a_{1k.23\ldots(k-1)(k+1)\ldots n}\, Y_k. \tag{6.1.28}$$

În scrierea (6.1.27) primul indice reprezintă variabila dependentă iar al doilea variabila independentă, iar împreună sînt cunoscuţi sub numele de prim-indici subscrişi, spre deosebire de cei de după punct care sînt numiţi indici secunzi (Yule, Kendall (1969)).

Un coeficient de regresie cu p indici secunzi se numeşte coeficient de regresie de ordinul p.

Dacă $p = 0$, avem un coeficient de regresie de ordin zero, cum este cazul pentru a_{12}, a_{21} şi alţii, pe care îi vom numi coeficienţi de regresie totală sau generală.

Observăm uşor că pentru $n = 2$ avem coeficientul de corelaţie totală

$$\rho_{12}^2 = a_{12}\ a_{21} \tag{6.1.29}$$

Definiţia 6.1.10. Se numeşte coeficient de corelaţie parţială de ordinul $(n-2)$ şi se notează cu $\rho^2_{12.34\ldots n}$ expresia

$$\rho^2_{12.34\ldots n} = a_{12.34\ldots n} \cdot a_{21.34\ldots n}. \tag{6.1.30}$$

Se poate arăta (Siegel (1956), Yule şi Kendall (1969), demonstraţia este destul de laborioasă) că

$$\rho^2_{12.34\ldots n} = \frac{(\rho_{12.34\ldots(n-1)} - \rho_{1n.34\ldots(n-1)} \cdot \rho_{2n.34\ldots(n-1)})^2}{(1 - \rho^2_{1n.34\ldots(n-1)})(1 - \rho^2_{2n.34\ldots(n-1)})} \tag{6.1.31}$$

unde $\rho_{12.34\ldots n}$ este definit de relaţia (6.1.30) din care pot fi deduşi $\rho_{1n.34\ldots(n-1)}$ şi ceilalţi, numiţi acum coeficienţi de corelaţie parţială de ordin inferior. Ca urmare, pentru $n = 3$, deducem

$$\rho^2_{12.3} = \frac{(\rho_{12} - \rho_{13} \cdot \rho_{23})^2}{(1 - \rho^2_{13})(1 - \rho^2_{23})} \tag{6.1.32}$$

rezultat ce poate fi obţinut şi direct, prin calcul, în virtutea relaţiilor (6.1.26) şi (6.1.29).

Scanned with OKEN Scanner

6.2. Măsuri ale concordanței

6.2.1. Coeficientul lui G. Fechner (1897)

Fie X și Y două variabile aleatoare discrete care iau n valori iar $M(X)$ și $M(Y)$ valorile medii ale acestora. Să notăm cu $\Delta x_k = x_k - M(X)$ și $\Delta y_k = y_k - M(Y)$ iar cu $\hat{X} = X - M(X)$ și $\hat{Y} = Y - M(Y)$ variabilele abateri de la medii corespunzătoare. Să notăm de asemenea cu $c(\hat{X}, \hat{Y})$ și $d(\hat{X}, \hat{Y})$ numărul de concordanțe și respectiv de discordanțe de semn ale perechilor $(\Delta x_k, \Delta y_k)$. Evident că $0 \leqslant c(\hat{X}, \hat{Y}), d(\hat{X}, \hat{Y}) \leqslant n$ și $c(\hat{X}, \hat{Y}) + d(\hat{X}, \hat{Y}) = n$.

Definiția 6.2.1. Se numește coeficient de concordanță a variabilelor X și Y în sensul lui Fechner mărimea

$$F(X, Y) = \frac{c(\hat{X}, \hat{Y}) - d(\hat{X}, \hat{Y})}{n} \qquad (6.2.1)$$

Într-o altă variantă, diferită ca expresie și rezultat, coeficientul de concordanță este definit considerînd

$\Delta x_k = x_k - x_{k-1}$, $\Delta y_k = y_k - y_{k-1}$. Deducem imediat

Propoziția 6.2.1. Pentru $F(X, Y)$ sînt adevărate următoarele

1) este definit pentru orice pereche (X, Y)

2) $F(X, Y) = F(Y, X)$

3) $-1 \leqslant F(X, Y) \leqslant 1$

4) $F(X, Y) = \pm 1$ dacă și numai dacă $d(\hat{X}, \hat{Y}) = 0$ sau $c(\hat{X}, \hat{Y}) = 0$

5) $F(X, Y) = 0$ dacă și numai dacă $c(\hat{X}, \hat{Y}) = d(\hat{X}, \hat{Y})$

Se spune, în sensul lui Fechner, că X și Y sînt total concordante sau total neconcordante după cum $d(\hat{X}, \hat{Y}) = 0$ sau $c(\hat{X}, \hat{Y}) = 0$ și că ele sînt egal concordante dacă $c(\hat{X}, \hat{Y}) = d(\hat{X}, \hat{Y})$.

Expresia (6.2.1) are neajunsul de a nu putea fi aplicată pentru variabile necantitative cînd $\Delta x_k = x_k - M(X)$ și $\Delta y_k = y_k - M(Y)$, motiv pentru care se alege varianta $\Delta x_k = x_k - x_{k-1}$ și $\Delta y_k = y_k - y_{k-1}$. Un alt neajuns a fost considerat faptul că se iau în seamă doar variațiile de semn, nu și valorile abaterilor de la medie. Pentru a corecta acest neajuns s-a propus coeficientul ponderat de concordanță. Să notăm

Scanned with OKEN Scanner

cu $I_1 = \{k | 1 \leqslant k \leqslant n,\ \Delta x_k \cdot \Delta y_k \geqslant 0\}$ și cu

$$I_2 = \{k | 1 \leqslant k \leqslant n,\ \Delta x_k \cdot \Delta y_k < 0\},\ I_1 \cap I_2 = \emptyset,$$

$$I_1 \cup I_2 = \{1,2,\ldots,n\}.$$

Coeficientul de concordanță are acum expresia

$$F_p(X,\ Y) = \frac{\sum\limits_{k \in I_1} \Delta x_k \cdot \Delta y_k - |\sum\limits_{k \in I_2} \Delta x_k \cdot \Delta y_k|}{\sum\limits_{k \in I_1} \Delta x_k \cdot \Delta y_k + |\sum\limits_{k \in I_2} \Delta x_k \cdot \Delta y_k|} \qquad (6.2.2)$$

cu aceleași proprietăți și semnificații ca și $F(X,\ Y)$. Deoarece atît $F(X,\ Y)$ cît și $F_p(X,\ Y)$ pot lua valorile ± 1 chiar dacă între variabilele X și Y nu există o dependență liniară, s-a propus un alt coeficient de concordanță care elimină și acest neajuns și anume, pentru X și Y uniform repartizate,

$$\hat{F}(X,Y) = \frac{\operatorname{cov}(\hat{X},\ \hat{Y})}{D(\hat{X}) \cdot D(\hat{Y})} \qquad (6.2.3)$$

care fiind un coeficient de corelație clasic are proprietățile acestuia.

6.2.2. Coeficientul lui C. Spearman (1904)

Fie X și Y două variabile aleatoare discrete, uniform repartizate, care descriu două caracteristici ale unei colectivități statistice. Dacă valorile acestora, cantitative sau calitative, sînt ordonate după un anumit criteriu, să notăm cu x_k și y_k rangul uneia dintre valori sau grupe valorice pentru X și Y. Gîndind pe X și Y din acest punct de vedere, putem spune că x_k și y_k sînt de fapt valorile lui X și Y. Evident că $M(X) = M(Y) = \dfrac{n+1}{2}$, în ipoteza că X și Y iau n valori, deci că avem n ranguri ale acestora. Să considerăm acum variabila $Z = X - Y$, ale cărei valori sînt $z_k = x_k - y_k$. Observăm că $M(Z) = 0$, fără ca z_k să fie zero. Dacă întîmplător $z_k = 0$, $1 \leqslant k \leqslant n$, s-ar spune că între X și Y există o asociere, sau o concordanță sau corespondență perfectă din punct de vedere al rangurilor și implicit a ceea ce reprezintă acestea. Să considerăm variabilele $\hat{X} = X - M(X)$ și $\hat{Y} = Y - M(Y)$.

Scanned with OKEN Scanner

Definiția 6.2.2. Se numește coeficient de corelație a rangurilor variabilelor aleatoare X și Y în sensul lui Spearman, expresia

$$s(X, Y) = \rho(\hat{X}, \hat{Y}) \tag{6.2.4}$$

unde $\rho(\hat{X}, \hat{Y})$ este coeficientul de corelație clasic.

Prin calcul direct deducem imediat că

$$s(X, Y) = 1 - \frac{6M(Z^2)}{n^2 - 1} = 1 - \frac{6}{n(n^2 - 1)} \sum_{k=1}^{n} z_k^2 \tag{6.2.5}$$

forme întîlnite mai des în practică.

Ca urmare a relației (6.2.4) deducem cu ușurință

Propoziția 6.2.2. Pentru $s(X, Y)$ sînt adevărate următoarele

1) este definit pentru orice pereche (X, Y)

2) $s(Y, X) = s(X, Y)$

3) $-1 \leqslant s(X, Y) \leqslant 1$

4) $s(X, Y) = \pm 1$ dacă și numai dacă corelația rangurilor este perfectă (directă cînd $s(X, Y) = 1$ și inversă cînd $s(X, Y) = -1$)

5) $s(X, Y) = 0$ cînd X și Y sînt necorelate din punct de vedere al rangurilor.

La cei peste 80 de ani ai săi, coeficientul lui Spearman rămîne încă un instrument util în practica statistică.

6.2.3. Coeficientul lui M. G. Kendall (1948)

Pe linia considerațiilor legate de coeficientul lui Spearman, Kendall a dat o altă expresie pentru un astfel de indicator. El ordonează rangurile lui X ca să apară natural și pentru că aceasta produce o schimbare a rangurilor lui Y introduce variabila aleatoare Z ale cărei valori z_k reprezintă diferența dintre numărul de ranguri superioare lui y_k și numărul de ranguri inferioare lui y_k dar nu inferioare lui y_{k-1}. Se poate observa că $\sum_{k=1}^{n} z_k \leqslant \frac{n(n-1)}{2}$, egalitatea avînd loc în cazul în care y_k apar în ordinea naturală $1, 2, \ldots, n$.

Definiția 6.2.3. Se numește coeficient de corelație a rangurilor pentru variabilele X și Y în sensul lui Kendall,

Scanned with OKEN Scanner

raportul

$$K(X, Y) = \frac{2}{n(n-1)} \sum_{k=1}^{n} z_k \qquad (6.2.6)$$

Se poate observa cu ușurință

Propoziția 6.2.3. Pentru $K(X, Y)$ sînt adevărate următoarele

1) este definit pentru orice cuplu (X, Y)
2) $K(X, Y) \neq K(Y, X)$
3) $-1 \leqslant K(X, Y) \leqslant 1$
4) $K(X, Y) = \pm 1$ dacă și numai dacă între X și Y există o concordanță deplină din punct de vedere al rangurilor sau o discordanță totală
5) $K(X, Y) = 0$ dacă X și Y sînt egal neconcordante din punct de vedere al rangurilor.

Cu toate că nu are o arie de aplicabilitate foarte mare, coeficientul lui Kendall rămîne, alături de coeficientul lui Spearman, ca un instrument util al practicii statistice clasice.

6.3. Media pătratului de contingență a lui K. Pearson (1901)

Deoarece, după cum se vede imediat, coeficientul de corelație clasic depinde de valorile pe care variabilele aleatoare le iau, aplicarea lui în cazul unor variabile necantitative devine dificilă, sau chiar imposibilă. Preocupat de acest fapt, K. Pearson a introdus în 1901 o altă măsură a dependenței care nu mai depinde decît de repartițiile variabilelor aleatoare în discuție. Vom prezenta acest rezultat urmîndu-l pe A. Rényi (1959).

Fie $\{\Omega, \mathcal{K}, P\}$ un cîmp de probabilitate și X_1 și X_2 două variabile aleatoare reale definite pe Ω. Să notăm cu P_{X_1} o măsură definită pentru orice mulțime boreliană $A \subseteq \mathbf{R}$ prin relația

$$P_{X_1}(A) = P(X_1 \in A) = P[X_1^{-1}(A)] \qquad (6.3.1)$$

unde

$$X_1^{-1}(A) = \{\omega \in \Omega / X_1(\omega) \in A\} \qquad (6.3.2)$$

Scanned with OKEN Scanner

Să notăm de asemenea cu P_{X_1,X_2} o măsură definită pentru orice mulțime boreliană $C \subseteq \mathbf{R}^2$ prin relația

$$P_{X_1,X_2}(C) = P[(X_1, X_2) \in C] = P[\vec{\xi}^{-1}(C)] \qquad (6.3.3)$$

unde $\vec{\xi}$ este vectorul aleator (X_1, X_2), iar $\vec{\xi}^{-1}(C)$ înseamnă

$$\vec{\xi}^{-1}(C) = \{\omega \in \Omega / \xi(\omega) = [X_1(\omega), X_2(\omega)] \in C\} \qquad (6.3.4)$$

și cu $P_{X_1 * X_2}$ o măsură definită pe mulțimi boreliene din plan prin relația

$$P_{X_1 * X_2}(A \times B) = P_{X_1}(A) \cdot P_{X_2}(B) \qquad (6.3.5)$$

unde

$$A \times B = \{(x_1, x_2) \in \mathbf{R}^2 / x_1 \in A \text{ și } x_2 \in B\} \qquad (6.3.6)$$

Măsura $P_{X_1 * X_2}$ definită de relația (6.3.5) se numește produs direct al măsurilor P_{X_1} și P_{X_2} definite de relația (6.3.1) și se spune că dependența variabilelor X_1 și X_2 este regulată dacă repartiția comună P_{X_1,X_2} este absolut continuă în raport cu produsul direct $P_{X_1 * X_2}$ al repartițiilor individuale P_{X_1} și P_{X_2} (Rényi (1959)).

Conform teoremei Radon-Nicodym, dacă dependența dintre X_1 și X_2 este regulată, atunci există o funcție măsurabilă Borel, $\varphi(x_1, x_2)$, unică aproape peste tot, astfel încît pentru orice mulțime boreliană $C \subseteq \mathbf{R}^2$ să avem relația

$$P_{X_1,X_2}(C) = \iint_C \varphi(x_1, x_2)\, \mathrm{d}P_{X_1 * X_2} \qquad (6.3.7)$$

unde funcția $\varphi(x_1, x_2)$, denumită și derivata Radon-Nicodym, are expresia

$$\varphi(x_1, x_2) = \frac{\mathrm{d}^2 P_{X_1, X_2}}{\mathrm{d}P_{X_1 * X_2}} = \frac{\mathrm{d}^2 P_{X_1, X_2}}{\mathrm{d}P_{X_1} \cdot \mathrm{d}P_{X_2}} \qquad (6.3.8)$$

Definiția 6.3.1. Cantitatea

$$C^2(X_1, X_2) = \iint_{\mathbf{R}^2} [\varphi(x_1, x_2) - 1]^2\, \mathrm{d}P_{X_1 * X_2} \qquad (6.3.9)$$

Scanned with OKEN Scanner

se numește, în sensul lui K. Pearson, medie a pătratului de contingență al variabilelor aleatoare X_1 și X_2.

Mărimea (6.3.9) poate fi definită atît pentru variabile continue cît și pentru variabile discrete. Dacă $f(x_1, x_2)$ este densitatea de repartiție a cuplului (X_1, X_2) iar $g(x_1)$ și $h(x_2)$ densitățile de repartiție individuale ale lui X_1 și X_2, atunci expresia (6.3.9) devine

$$C^2(X_1, X_2) = \iint_{R^2} \left[\frac{f(x_1, x_2)}{g(x_1)h(x_2)} - 1 \right]^2 g(x_1)\, h(x_2) \mathrm{d}x_1\, \mathrm{d}x_2 =$$

$$= \iint_{R^2} \frac{f^2(x_1, x_2)}{g(x_1)h(x_2)}\, \mathrm{d}x_1\, \mathrm{d}x_2 - 1 \qquad (6.3.10)$$

în cazul continuu, sau

$$C^2(X_1, X_2) = \sum_{x_1} \sum_{x_2} \left[\frac{f(x_1, x_2)}{g(x_1)h(x_2)} - 1 \right]^2 g(x_1)\, h(x_2) =$$

$$= \sum_{x_1} \sum_{x_2} \frac{f^2(x_1, x_2)}{g(x_1)\, h(x_2)} - 1 \qquad (6.3.11)$$

în cazul discret și care este forma clasică dată de K. Pearson.

Se poate demonstra (vezi și (Mihoc (1966), Rényi (1959))

Propoziția 6.3.1. Pentru media patratului de contingență $C^2(X_1, X_2)$ sînt adevărate următoarele :

1) este definită pentru orice cuplu aleator (X_1, X_2) continuu sau discret
2) este simetrică în X_1 și X_2 adică, $C^2(X_1, X_2) = C^2(X_2, X_1)$
3) $0 \leqslant C^2(X_1, X_2) < \infty$ pentru orice cuplu (X_1, X_2)
4) $C^2(X_1, X_2) = 0$ dacă și numai dacă X_1 și X_2 sînt independente probabilistic
5) este invariantă în raport cu orice transformare reală biunivocă a variabilelor X_1 și X_2
6) dacă (X_1, X_2) este repartizat normal, atunci

$$C^2(X_1, X_2) = \frac{\rho^2(X_1, X_2)}{1 - \rho^2(X_1, X_2)} \qquad (6.3.12)$$

7) dacă X_1 și X_2 sînt variabile discrete luînd m și respectiv n valori, atunci

$$C^2(X_1, X_2) \leqslant \sqrt{(m-1)(n-1)} \qquad (6.3.13)$$

cu egalitate dacă $X_1 = \psi_2(X_2)$ sau $X_2 = \psi_1(X_1)$, ψ_1 și ψ_2 fiind funcții reale definite pe axa reală în mod biunivoc.

Ca urmare a proprietăților 3) și 7) putem da

Definiția 6.3.2. Se numesc rapoarte de contingență corespunzătoare vectorului aleator (X_1, X_2) sau coeficienți de contingență ai lui Pearson, mărimile

$$C_1^2(X_1, X_2) = \frac{C^2(X_1, X_2)}{1 + C^2(X_1, X_2)} \quad \text{pentru orice } (X_1, X_2) \tag{6.3.14}$$

$$C_2^2(X_1, X_2) = \frac{C^2(X_1, X_2)}{\sqrt{(m-1)(n-1)}} \quad \text{pentru } X_1 \text{ și } X_2 \text{ discrete cu } m \text{ și respectiv } n \text{ valori.} \tag{6.3.15}$$

Astfel, proprietatea 3) poate fi reformulată ca

3') $0 \leqslant C_1^2(X_1, X_2) < 1$ și $0 \leqslant C_2^2(X_1, X_2) \leqslant 1$.

Deducem de asemenea că dacă (X_1, X_2) este repartizat normal atunci

$$C_1^2(X_1 X_2) = \rho^2(X_1, X_2). \tag{6.3.16}$$

Să remarcăm că spre deosebire de coeficientul de corelație $\rho(X_1, X_2)$, media pătratului de contingență și coeficienții corespunzători reprezintă măsuri ale dependenței probabilistice dintre două repartiții statistice. Din acest motiv ele pot fi considerate ca o extensie a măsurii dependenței de la două variabile aleatoare la două familii de variabile aleatoare. Se observă de asemenea aplicabilitatea acestora atît pentru variabile aleatoare cantitative cît și pentru cele calitative. Cu toate acestea, postulatele lui Rényi nu sînt îndeplinite în întregime și doar $C_1^2(X_1, X_2)$ este mai apropiat de modelul de „bună măsură a dependenței" în sensul lui Rényi.

6.4. Corelația maximală Gebelein-Kramer (1941)

Preocupat de îmbunătățirea măsurilor de corelație Galton-Pearson, H. Gebelein a definit în 1941 un nou indicator în acest sens, căruia H. P. Kramer (1948) și A. Rényi (1959)

Scanned with OKEN Scanner

i-au adus unele îmbunătăţiri şi o anumită rigoare în prezentare. De altfel Rényi a studiat şi utilizat acest indicator pentru a ilustra verificarea postulatelor unei bune măsuri a dependenţei, formulate de el în 1959.

Fie X_1 şi X_2 două variabile aleatoare a căror repartiţie comună este $p(x_1, x_2)$, iar repartiţiile individuale sînt $p(x_1)$ şi $p(x_2)$, definite respectiv pe $D^2 \subseteq \mathbf{R}^2$ şi $D_1, D_2 \subseteq \mathbf{R}$. Presupunem cunoscute valorile medii condiţionate şi dispersiile condiţionate corespunzătoare variabilelor X_1 şi X_2 concepte deja utilizate în paragrafele precedente. Fie acum două funcţii arbitrare $f, g : \mathbf{R} \to \mathbf{R}$ şi să considerăm expresia

$$\mathcal{G}^2[X_1, X_2] = \frac{[M[f(X_1)\, g(X_2)]]^2}{M[f^2(X_1)M[g^2(X_2)]}, \tag{6.4.1}$$

care în cazul discret are forma

$$\mathcal{G}^2[X_1, X_2] = \frac{[\sum\limits_{x_1} \sum\limits_{x_2} p(x_1, x_2)\, f(x_1)\, g(x_2)]^2}{[\sum\limits_{x_1} p(x_1) f^2(x_1)]\,[\sum\limits_{x_2} p(x_2)\, g^2(x_2)]}, \tag{6.4.2}$$

iar în cazul continuu devine

$$\mathcal{G}^2[X_1, X_2] = \frac{\left[\iint\limits_D p(x_1, x_2) f(x_1)\, g(x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2\right]^2}{\left[\int\limits_{D_1} p(x_1) f^2(x_1)\, \mathrm{d}x_1\right]\left[\int\limits_{D_2} p(x_2) g^2(x_2)\, \mathrm{d}x_2\right]}. \tag{6.4.3}$$

Ca urmare a inegalităţii lui Schwarz-Cauchy-Buniakovsky, din (6.4.1) rezultă că pentru orice (X_1, X_2) şi (f, g) pentru care relaţia are sens, avem

$$\mathcal{G}^2[X_1, X_2] \leqslant 1 \tag{6.4.4}$$

cu egalitate dacă şi numai dacă $f(x_1) = \alpha g(x_2)$ pentru orice x_1 şi x_2, ceea ce înseamnă că X_1 şi X_2 sînt funcţional dependente.

Definiţia 6.4.1. Fie f_0 şi $g_0 : \mathbf{R} \to \mathbf{R}$ cu proprietatea

$$\mathcal{G}_0^2[X_1, X_2] = \frac{[M[f_0(X_1)\, g_0(X_2)]]^2}{M[f_0^2(X_1)]\, M[g_0^2(X_2)]} =$$

$$= \sup_{(f,g)} \frac{[M[f(X_1)\, g(X_2)]]^2}{M[f^2(X_1)]\, M[g^2(X_2)]}. \tag{6.4.5}$$

Scanned with OKEN Scanner

Dacă acest supremum există, atunci valoarea sa $\mathcal{G}_0^2[X_1, X_2]$ se numește corelație maximală, în sensul lui Gebelein și Kramer, a variabilelor aleatoare X_1 și X_2.

Remarcăm cu ușurință că dacă f și g sînt astfel definite încît $M(f(X_1)) = M(g(X_2)) = 0$, atunci

$$\mathcal{G}^2[X_1,\ X_2] = \rho^2[f(X_1),\ g(X_2)] \tag{6.4.6}$$

unde ρ este coeficientul de corelație clasic. De altfel, relația (6.4.1), dacă ținem cont și de (6.4.4), poate defini și ea un indicator al dependenței variabilelor X_1 și X_2 care însă nu satisface ca naturală axioma independenței.

Să presupunem acum, fără a diminua generalitatea problemei, că X_1, X_2, f și g sînt definite astfel încît $M(X_1) = M(X_2) = 0$, $D(X_1) = D(X_2) = 1$, $M[f(X_1)] = M[g(X_2)] = 0$ și $D[f(X_1)] = D[g(X_2)] = 1$. Să considerăm atunci mărimea (6.4.6). Ținînd cont și de relația (6.1.12) putem scrie că

$$M^2[f(X_1)g(X_2)] = M^2\{f(X_1)M[g(X_2)/X_1]\} =$$

$$= M^2\{M[f(X_1)/X_2]g(X_2)\} \tag{6.4.7}$$

și aplicînd inegalitatea lui Schwarz, în fiecare caz, deducem

$$M^2[f(X_1)g(X_2)] \leqslant M[M^2[f(X_1)/X_2]] \cdot M[g^2(X_2)] \leqslant$$

$$\leqslant M[M^2[f(X_1)/X_2]] \tag{6.4.8}$$

cu egalitate dacă și numai dacă

$$g_0(X_2) = a\,M[f(X_1)/X_2] + b \tag{6.4.9}$$

Cum $M[g(X_2)] = 0$ și $D[g(X_2)] = 1$ rezultă că

$$g_0(X_2) = \frac{M[f_0(X_1)/X_2]}{D[M\ [f_0(X_1)/X_2]} \tag{6.4.10}$$

și în mod analog

$$f_0(X_1) = \frac{M[g_0(X_2)/X_1]}{D[M[g_0(X_2)/X_1]]} \tag{6.4.11}$$

adică funcțiile $f_0(X_1)$ și $g_0(X_2)$ care maximizează $\mathcal{G}^2(X_1, X_2)$

Scanned with OKEN Scanner

Fie $\lambda^2 = \max_{(f,g)} \mathcal{G}^2(X_1, X_2)$. Dacă (6.4.8) este egalitate, deci λ^2 există, atunci putem scrie că

$$\lambda f_0(X_1) = M[g_0(X_2)/X_1] \text{ și } \lambda g_0(X_2) = M[f_0(X_1)/X_2] \tag{6.4.12}$$

Avînd în vedere cele de mai sus putem considera ca demonstrată

Propoziția 6.4.1. Dacă maximul expresiei (6.4.1) există, atunci funcțiile $f_0(x_1)$, și $g_0(x_2)$ care îl realizează sînt date de relațiile (6.4.10) și (6.4.11).

Din relația (6.4.10) observăm că dacă $f_0(x_1)$ este cunoscut, atunci putem obține pe $g_0(x_2)$ și analog din (6.4.11). De asemenea din (6.4.12) deducem că

$$\lambda^2 f_0(X_1) = M[M[f_0(X_1)/X_2]] \text{ sau}$$

$$\lambda^2 g_0(X_2) = M[M[g_0(X_2)/X_1]/X_2] \tag{6.4.13}$$

relații care constituie ecuațiile din care pot fi deduse $f_0(X_1)$ și $g_0(X_2)$.

Să presupunem că X_1 ia m valori, iar X_2 n valori și să notăm cu $p_{ij} = P(X_1 = x_1^i,\ X_2 = x_2^j)$, $q_i = P(X_1 = x_1^i)$, $r_j = P(X_2 = x_2^j)$, $f_i = f_0(x_1^i)$ și $g_j = g_0(x_2^j)$, $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$. Atunci prima din relațiile (6.4.13) devine

$$\lambda^2 f_i = \sum_{j=1}^{n} \left[\sum_{k=1}^{m} f_k \frac{p_{kj}}{r_j} \right] \frac{p_{ij}}{q_i}, \quad 1 \leqslant i \leqslant m \tag{6.4.14}$$

sau

$$\lambda^2 f_i = \sum_{k=1}^{m} \left[\sum_{j=1}^{n} \frac{p_{ij}\, p_{kj}}{q_i\, r_j} \right] f_k, \quad 1 \leqslant i \leqslant m \tag{6.4.15}$$

sau încă

$$\lambda^2 f_i = \sum_{k=1}^{m} a_{ik} f_k, \quad 1 \leqslant i \leqslant m \tag{6.4.16}$$

unde

$$a_{ik} = \sum_{j=1}^{n} \frac{p_{ij}\, p_{kj}}{q_i\, r_j}, \quad 1 \leqslant i,\ k \leqslant m \tag{6.4.17}$$

sau matriceal

$$Af = \lambda^2 f \tag{6.4.18}$$

Scanned with OKEN Scanner

unde $A = (a_{ik})$ și $f = (f_i)$. Relația (6.4.19) reprezintă un sistem de m ecuații cu m necunoscute f_i, care admite și soluție nenulă dacă și numai dacă

$$|A - \lambda^2 E| = 0 \tag{6.4.19}$$

Să remarcăm că A este o matrice stochastică ($a_{ik} \geqslant 0$, $\sum_{k=1}^{m} a_{ik} = 1$) și ca urmare a acestui fapt (vezi și Gh. Mihoc (1966)) ecuația (6.4.19) admite m rădăcini reale λ_i^2, $0 \leqslant \lambda_i^2 \leqslant 1$, $1 \leqslant i \leqslant m$. Cum λ^2 reprezintă valoarea maximă a expresiei (6.4.1) rezultă că vom alege dintre soluțiile λ_i^2 ale ecuației (6.4.19) pe cea mai mare.

Se poate observa imediat că relația (6.4.18) se obține și dacă am considera cea de a doua ecuație (6.4.13) Am demonstrat

Propoziția 6.4.2. Fie λ_0^2 cea mai mare dintre rădăcinile ecuației (6.4.19). Atunci

$$\mathcal{G}_{f_0}^2[X_1, X_2] = \lambda_0^2 \tag{6.4.20}$$

Ca urmare a propoziției 6.4.2 deducem imediat

1) expresiile analitice ale funcțiilor f_0 și g_0 care maximizează pe $\mathcal{G}^2(X_1, X_2)$ nu mai sînt necesare
2) corelația maximală este determinată numai cu ajutorul repartițiilor comună și individuale ale variabilelor X_1 și X_2, fără a lua în considerație și valorile acestora.
3) calculul efectiv al valorii corelației maximale dintre X_1 și X_2 constă din :

3.1) determinarea matricei $A = (a_{ik})$ dată de relația (6.4.17)

3.2) rezolvarea ecuației $|A - \lambda^2 E| = 0$

3.3) alegerea celei mai mari dintre rădăcinile ecuației $|A - \lambda^2 E| = 0$ care reprezintă valoarea corelației maximale căutate.

Propoziția 6.4.3. Dacă $\lambda_1^2 \leqslant \lambda_2^2 \leqslant \ldots \leqslant \lambda_{m-1}^2 \leqslant 1$ sînt rădăcinile ecuației (6.4.19) atunci între corelația maximală $\mathcal{G}_{f_0}^2(X_1, X_2)$ și media pătratului de contingență $C^2(X_1, X_2)$ are loc relația

$$\mathcal{G}_{f_0}^2(X_1, X_2) \geqslant \sqrt{(n-1)/(m-1)}\, C^2(X_1, X_2) \geqslant C^2(X_1, X_2) (n \geqslant m) \tag{6.4.21}$$

cu egalitate dacă și numai dacă A este matricea unitate.

Scanned with OKEN Scanner

Demonstrație. Dacă $\lambda_1^2 \leqslant \lambda_2^2 \leqslant \ldots \leqslant \lambda_{m-1}^2 \leqslant 1$ sînt rădăcinile ecuației (6.4.19) atunci

$$\sum_{i=1}^{m-1} \lambda_i^2 = \sum_{i=1}^{m} a_{ij} - 1 = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{p_{ij}^2}{q_i r_j} - 1 =$$

$$= \sqrt{(m-1)(n-1)}\, C^2(X_1, X_2) \qquad (6.4.22)$$

Ca urmare, cum $\lambda_0^2 = \lambda_{m-1}^2$

$$(m-1)\lambda_0^2 \geqslant \sqrt{(m-1)(n-1)}\; C^2(X_1, X_2) \qquad (6.4.23)$$

de unde rezultă (6.4.21). Relația (6.4.23) și (implicit (6.4.21)) devine egalitate dacă $\lambda_1^2 = \lambda_2^2 = \ldots = \lambda_{m-1}^2 = 1$, și $m = n$ ceea ce este posibil dacă și numai dacă $a_{ij} = 1$ și $a_{ij} = 0$ $(i \neq j)$.

Propoziția 6.4.4. Pentru corelația maximală $\mathcal{G}_0^2(X_1, X_2)$ sînt adevărate totdeauna următoarele afirmații

1) este definită pentru orice vector aleator (X_1, X_2)
2) este simetrică în X_1 și X_2, adică $\mathcal{G}_0^2(X_1, X_2) = \mathcal{G}_0^2(X_2, X_1)$
3) $0 \leqslant \mathcal{G}_0^2(X_1, X_2) \leqslant 1$
4) $\mathcal{G}_0^2(X_1, X_2) = 0$ dacă și numai dacă X_1 și X_2 sînt independente
5) $\mathcal{G}_0^2(X_1, X_2) = 1$ dacă și numai dacă X_1 și X_2 sînt funcțional dependente
6) este invariantă în raport cu orice transformare biunivocă a variabilelor X_1 și X_2, repartizate discret, adică

$$\mathcal{G}_0^2[\varphi_1(X_1), \varphi_2(X_2)] = \mathcal{G}_0^2[X_1, X_2] \qquad (6.4.24)$$

pentru orice $\varphi_1, \varphi_2 : \mathbf{R} \to \mathbf{R}$ definite în mod biunivoc;
7) dacă (X_1, X_2) este repartizat normal, atunci

$$\mathcal{G}_0^2(X_1, X_2) = \rho^2(X_1, X_2) \qquad (6.4.25)$$

Demonstrație. Proprietățile 1) și 2) sînt evidente, 3) rezultă din (6.4.4.), iar 6) din faptul că orice transformare biunivocă a unei variabile aleatoare discrete nu modifică repartiția probabilistică a acesteia. Dacă X_1 și X_2 sînt independente atunci $C^2(X_1, X_2) = 0$ și ca urmare a relației

Scanned with OKEN Scanner

(6.4.22) rezultă că $\sum_{i=1}^{m-1} \lambda_i^2 = 0$ ceea ce implică $\lambda_1^2 = \lambda_2^2 = \ldots = \lambda_{m-1}^2 = 0$ și deci $\mathcal{G}_{f_0}^2(X_1, X_2) = 0$. Reciproc, dacă $\mathcal{G}_{f_0}^2(X_1, X_2) = 0$, atunci $\lambda_{m-1}^2 = 0$ și ca urmare și celelalte rădăcini ale ecuației (6.4.19) sînt nule. De aici rezultă $C^2(X_1, X_2) = 0$ ceea ce înseamnă $p_{ij} = q_i r_j$ pentru orice (i, j), $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$, adică X_1 și X_2 sînt independente, și ca urmare, 4) este demonstrată. Dacă X_1 și X_2 sînt dependente funcțional, atunci coeficientul de contingență al lui Pearson $C_2^2(X_1, X_2) = 1$ și ca urmare a lui (6.4.2) rezultă $\mathcal{G}_{f_0}^2(X_1, X_2) \geqslant 1$. Ținînd cont și de (6.4.4) rezultă $\mathcal{G}_{f_0}^2(X_1, X_2) = 1$. Reciproc, dacă $\mathcal{G}_{f_0}^2(X_1, X_2) = 1$, atunci, din relația

$$f_i g_j = \frac{1}{2} [f_i^2 + g_j^2 - (f_i - g_j)^2] \tag{6.4.26}$$

rezultă

$$\mathcal{G}_{f_0}^2(X_1, X_2) = \sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij} f_i g_j = 1 -$$

$$- \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij}(f_i - g_j)^2 \qquad (p_{ij} \neq 0) \tag{6.4.27}$$

și ca urmare

$$\sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij}(f_i - g_j)^2 = 0, \tag{6.4.28}$$

relație care are loc dacă și numai dacă $f_i = g_j$ pentru orice i și j, ceea ce înseamnă $f_0(x_1^i) = g_0(x_2^j)$ pentru orice i și j, deci X_1 și X_2 sînt dependente funcțional (cu excepția cazului trivial $f(X_1) = g(X_2) = \text{constant}$). Prin urmare este demonstrată și 6). Ținînd cont și de (6.1.13), dacă X_1 și X_2 sînt normal repartizate, atunci (vezi și Rényi (1959))

$$f_0(X_1) = \frac{M(X_2/X_1)}{D[M(X_2/X_1)]} = \frac{X_1}{D(X_1)} \quad \text{și} \quad g(X_2) =$$

$$= \frac{M(X_1/X_2)}{D[M(X_1/X_2)]} = \frac{X_2}{D(X_2)} \tag{6.4.29}$$

și ca urmare $\mathcal{G}^2(X_1, X_2) = \rho^2[f(X_1), f(X_2)] = \rho^2$, deci proprietatea 7). Ținînd cont de cele de mai sus, deducem cu ușurință că $\mathcal{G}_{f_0}^2(X_1, X_2)$ constituie o bună măsură a dependenței în sensul lui Rényi (1959). În ciuda acestui fapt remarcabil,

Scanned with OKEN Scanner

$\mathcal{G}_0^2(X_1, X_2)$ nu s-a impus în practică probabil pentru că utilizarea lui este mai incomodă, necesitînd calcule mai laborioase decît coeficientul de corelaţie clasic $\rho(X_1, X_2)$.

6.5. Coeficientul asimptotic de corelaţie Onicescu–Săcuiu (1964)

Fie X_1 şi X_2 două variabile aleatoare arbitrare definite pe cîmpul de probabilitate $\{\Omega, \mathcal{K}, P\}$. Să considerăm mărimea

$$r_{a_1,a_2}(X_1, X_2) = \frac{\int_D X_1(\omega)\, X_2(\omega)\, \mathrm{d}P(\omega)}{\left(\int_{D_1} |X_1(\omega)|^2\, \mathrm{d}P(\omega)\right)^{1/2} \left(\int_{D_2} |X_2(\omega)|^2 \mathrm{d}P(\omega)\right)^{1/2}} \tag{6.5.1}$$

unde $D_i = \{\omega \,|X_i(\omega)| < a_i\}$, $i = 1, 2$, $D = D_1 \cap D_2$, $a_1, a_2 \in \in (0, \infty)$, familiile $\{D_i(a_i)\}_{a_i>0}$ $i = 1, 2$ fiind crescătoare şi cu limita Ω. Ca urmare a inegalităţii lui Schwarz, avem

$$|r_{a_1,a_2}(X_1, X_2)| \leqslant 1 \tag{6.5.2}$$

Să notăm cu

$$\underline{r}(X_1, X_2) = \varliminf_{a_1,a_2\to\infty} r_{a_1,a_2}(X_1, X_2\,; \bar{r}(X_1, X_2) =$$

$$= \varlimsup_{a_1,a_2\to\infty} r_{a_1,a_2}(X_1, X_2) \tag{6.5.3}$$

respectiv limitele inferioară şi superioară ale şirului

$$\{r_{a_1,a_2}(X_1, X_2)\}_{\substack{a_1>0\\ a_2>0}}$$

Definiţia 6.5.1. Dacă există, mărimea

$$r(X_1, X_2) = \underline{r}(X_1, X_2) = \bar{r}(X_1, X_2) \tag{6.5.4}$$

se numeşte coeficient asimptotic de corelaţie în sensul lui Onicescu şi Săcuiu. Din modul în care este definit rezultă că $r(X_1, X_2)$ poate exista chiar şi pentru variabile aleatoare ale căror dispersii nu sînt finite. Fiind o extensie a coeficientului de corelaţie clasic, păstrează proprietăţile acestuia din pro-

Scanned with OKEN Scanner

poziția 6.1.1 cu adaptarea corespunzătoare a notației. Se poate remarca de asemenea că dacă $D^2(X_1) < \infty$ și $D^2(X_2) < \infty$, atunci coeficientul asimptotic și cel clasic de corelație coincid.

6.6. Coeficientul de corelație globală al lui I. Săcuiu (1970)

În 1969 I. Săcuiu a dat o extensie a coeficientului de corelație clasic considerînd trei variabile aleatoare, rezultat extins la rîndul lui în 1970 pentru n variabile aleatoare $X_1, X_2, \ldots X_n$ ale căror momente de diverse ordine există.

Definiția 6.6.1. Se numește covarianță a variabilelor aleatoare $X_1, \ldots, X_n$ în sensul lui I. Săcuiu, expresia

$$\operatorname{cov}(X_1, X_2, \ldots, X_n) = M\left[\prod_{j=1}^{n}(X_j - m_j)\right] \qquad (6.6.1)$$

unde $m = (m_1, \ldots, m_n)$ este vectorul valorilor medii ale variabilelor $X_1, \ldots, X_n$.

Definiția 6.6.2. Se numește coeficient de corelație globală al variabilelor aleatoare $X_1, X_2, \ldots, X_n$ în sensul lui I. Săcuiu, raportul

$$S(X_1, \ldots, X_n) = \frac{M\left[\prod_{j=1}^{n}(X_j - m_j)\right]}{\left[\prod_{j=1}^{n} M(|X_j - m_j|^n)\right]^{1/n}} \qquad (6.6.2)$$

Se observă că pentru $n = 2$

$$S(X_1, X_2) = \rho(X_1, X_2) \qquad (6.6.3)$$

Definiția 6.6.3. Se spune că variabilele aleatoare $X_1, X_2, \ldots, X_n$ sînt total necorelate dacă $\operatorname{cov}(X_1, X_2, \ldots, X_n) = 0$.

Rezultă imediat

Propoziția 6.6.1. Dacă $X_1, X_2, \ldots, X_n$ sînt total independente atunci

$$S(X_1, X_2, \ldots, X_n) = 0.$$

Dezvoltînd în (6.6.2) se poate demonstra (Săcuiu (1970))

Scanned with OKEN Scanner

Propoziția 6.6.2. Dacă $X_1, X_2, \ldots, X_n$ sînt independente s cîte $s(s < n)$, atunci

$$S(X_1, \ldots, X_n) =$$

$$= \left[\prod_{j=1}^{n} M(|X_j - m_j|^n)\right]^{-\frac{1}{n}} \left\{M\left(\prod_{j=1}^{n} X_j\right) - \sum_{j=1}^{n} m_j M\left(\sum_{k=1}^{n(j)} X_k\right) + \ldots +\right.$$

$$+ (-1)^{n-s-1} \sum_{j_1, j_2, \ldots, j_{n-s-1}} m_{j_1} m_{j_2} \ldots$$

$$\ldots m_{j_{n-s-1}} M\left[\prod_{k=1}^{n(j_1 j_2, \ldots, j_{n-s-1})} X_k\right] +$$

$$\left. + (-1)^n (1 - C_n^1 + C_n^2 + \ldots + (-1)^s C_n^s) \prod_{j=1}^{n} M_j. \right. \qquad (6.6.4)$$

Ținînd cont de cele de mai sus se poate demonstra (vezi și Săcuiu (1970))

Propoziția 6.6.3. Pentru coeficientul de corelație globală $S(X_1, \ldots, X_n)$ sînt adevărate următoarele

1) este definit pentru orice vector aleator $(X_1, X_2, \ldots, X_n)$ ale cărui componente nu sînt constante;

2) este simetric în raport cu orice permutare a componentelor vectorului

3) $S^2(X_1, X_2, \ldots, X_n) \leqslant 1$

4) dacă $X_1, \ldots, X_n$ sînt total independente, atunci $S(X_1, \ldots, X_n) = 0$ nu și reciproc

5) $S^2(X_1, \ldots, X_n) = 1$ dacă și numai dacă $|X_i - m_i|^n = |\alpha_i| |X_1 - m_1|^n$ pentru orice $2 \leqslant i \leqslant n$, unde α_i sînt constante, aproape sigur

6) este invariant în raport cu orice transformare reală biunivocă a variabilelor aleatoare $X_1, X_2, \ldots, X_n$ repartizate discret.

Este evident că fiind construit ca o extensie a coeficientului de corelație clasic, coeficientul de corelație globală păstrează unele din dezavantajele acestuia. Cu toate acestea este un mare pas spre studierea dependenței globale dintre mai mult de două variabile aleatoare.

Scanned with OKEN Scanner

7. Măsuri informaționale

Vom prezenta sub acest titlu rezultatele cele mai importante privind măsurarea dependenței probabilistice, care au fost formulate după apariția teoriei informației și care au la bază ideile și conceptele acesteia.

7.1. Coeficientul logaritmic de corelație M. C. Camargo–I. Medina (1955)

Fie X o variabilă aleatoare discretă avînd repartiția $p_i \geqslant 0, \sum_{i=1}^{n} p_i = 1$

Definiția 7.1.1. Se numește certitudine sau siguranță a distribuției probabilistice discrete $p_i \geqslant 0, \sum_{i=1}^{n} p_i = 1$, în sensul lui Camargo și Medina, cantitatea

$$C(X) = \log n + \sum_{i=1}^{n} p_i \log p_i \tag{7.1.1}$$

Să observăm că folosind entropia Shannon și rescriind relația (7.1.1) sub forma

$$C(X) = -\sum_{i=1}^{n} \frac{1}{n} \log n - \left(-\sum_{i=1}^{n} p_i \log p_i \right) \tag{7.1.2}$$

atunci certitudinea sau siguranța unei repartiții discrete apare ca variație a cantității de informație cînd trecem de la repartiția uniformă la repartiția arbitrară $p_i \geqslant 0, \sum_{i=1}^{n} p_i = 1$. Ea apare totodată și ca diferență între măsura lui Hartley și măsura lui Shannon pentru cantitatea de informație furnizată de un experiment cu n rezultate posibile.

Să considerăm acum două variabile aleatoare discrete (X_1, X_2) a căror repartiție comună este $p(x_1, x_2)$ și ale căror repartiții individuale sînt $p(x_1)$ și $p(x_2)$. Avînd în vedere inegalitatea

$$\sum_{x_1} \sum_{x_2} p(x_1)\, p(x_2) \log p(x_1)p(x_2) \leqslant \sum_{x_1} \sum_{x_2} p(x_1, x_2) \log p(x_1, x_2) \tag{7.1.3}$$

Scanned with OKEN Scanner

care devine egalitate dacă și numai dacă X_1 și X_2 sînt independente să considerăm expresia

$$r_0(X_1, X_2) = \sum_{x_1} \sum_{x_2} [p(x_1, x_2) \log p(x_1, x_2) - \\ - p(x_1)\, p(x_2) \log p(x_1)\, p(x_2)] \qquad (7.1.4)$$

Definiția 7.1.2. Cantitatea $r_0(X_1, X_2)$ definită de relația (7.1.4) se numește indice logaritmic de corelație, în sensul lui Camargo și Medina.

Dacă X_1 și X_2 iau respectiv n_1 și n_2 valori, atunci, notînd cu

$$\Delta_{X_i} = \log n_i + \sum_{x_i} p(x_i) \log p(x_i), \qquad i = 1,\ 2 \qquad (7.1.5)$$

$$\Delta_{X_1 X_2} = \log n_1 n_2 + \sum_{x_1} \sum_{x_2} p(x_1, x_2) \log p(x_1, x_2) \qquad (7.1.6)$$

certitudinea corespunzătoare lui X_1, X_2 și (X_1, X_2), putem scrie că

$$r_0(X_1, X_2) = \Delta_{X_1 X_2} - (\Delta_{X_1} + \Delta_{X_2}) \qquad (7.1.7)$$

și ca urmare M. Castans Camargo și Medina e Isabel interpretează pe r_0 ca fiind variația certitudinii distribuției discrete a vectorului aleator (X_1, X_2) sau ca un spor de informație cînd se trece de la ipoteza de independență la cea de dependență totală a variabilelor X_1 și X_2.

Definiția 7.1.3. Se numește coeficient logaritmic de corelație al variabilelor aleatoare discrete X_1 și X_2, în sensul lui Camargo și Medina, raportul

$$R_0(X_1, X_2) = \frac{-2r_0(X_1, X_2)}{\sum_{x_1} \sum_{x_2} p(x_1)\, p(x_2) \log [p(x_1)\, p(x_2)]}. \qquad (7.1.8)$$

Se poate deduce cu ușurință

Propoziția 7.1.1. Pentru coeficientul logaritmic de corelație $R_0(X_1, X_2)$ sînt adevărate următoarele afirmații

1) este definit pentru orice vector aleator discret neconstant
2) este simetric în raport cu X_1 și X_2, adică,

$$R_0(X_1, X_2) = R_0(X_2, X_1)$$

Scanned with OKEN Scanner

3) $0 \leqslant R_0(X_1, X_2) \leqslant 1$

4) $R_0(X_1, X_2) = 0$ dacă și numai dacă X_1 și X_2 sînt independente

5) $R_0(X_1, X_2) = 1$ dacă și numai dacă X_1 și X_2 sînt dependente funcțional

6) este invariant în raport cu orice transformare reală biunivocă a variabilelor X_1 și X_2, adică, pentru orice $\varphi_1\ \varphi_2 : \mathbf{R} \to \mathbf{R}$, biunivoce, $R_0[\varphi_1(X_1), \varphi_2(X_2)] = R_0(X_1, X_2)$.

În mod analog cazului discret se pot defini mărimile $r_0(X_1, X_2)$ și $R_0(X_1, X_2)$ și pentru variabile aleatoare continue, înlocuind simbolul sumă cu cel de integrală. Prin calcul direct obținem imediat,

Propoziția 7.1.2. Dacă (X_1, X_2) este repartizat normal, atunci

$$r_0(X_1, X_2) = -\frac{1}{2} \log [1 - \rho^2(X_1, X_2)] \qquad (7.1.9)$$

unde $\rho(X_1, X_2)$ este coeficientul de corelație clasic.

Să remarcăm că exceptînd cel de-al șaptelea postulat al lui Rényi, celelalte șase postulate ale unei bune măsuri a dependenței în sensul lui Rényi sînt îndeplinite de către coeficientul logaritmic de corelație $R_0(X_1, X_2)$.

7.2. Coeficientul de corelație informațională al lui E. H. Linfoot (1957)

Preluînd rezultatele spaniolilor M. Castans Camargo și Medina e Isabel, Linfoot stabilește în 1957 o legătură între indicele logaritmic de corelație și coeficientul clasic de corelație și definește astfel o nouă măsură a dependenței.

Propoziția 7.2.1. Dacă densitatea de repartiție a vectorului (X_1, X_2) este

$$p(x_1, x_2) = \frac{1}{2\pi} \sqrt{a_1 a_2 - h^2} \exp\left\{-\frac{1}{2}(a_1 x_1^2 + 2h x_1 x_2 + a_2 x_2^2)\right\},$$

$$a_1 > 0,\ a_2 > 0,\ a_1 a_2 > h^2 \qquad (7.2.1)$$

Scanned with OKEN Scanner

atunci între indicele logaritmic de corelaţie $r_0(X_1, X_2)$ şi coeficientul de corelaţie clasic $\rho(X_1, X_2)$ are loc relaţia

$$\rho(X_1, X_2) = 1 - \exp\{-2\ r_0(X_1, X_2)\} \tag{7.2.2}$$

Demonstraţie. Calculînd densităţile de repartiţie individuale obţinem

$$p(x_1) = \sqrt{\frac{a_1 a_2 - h^2}{2 a_2 \pi}} \exp\left\{-\frac{a_1 a_2 - h^2}{a_2} x_1^2\right\} \tag{7.2.3}$$

şi analog pentru $p(x_2)$. Ca urmare, prin calcul direct, obţinem

$$\rho(X_1, X_2) = \frac{-h}{\sqrt{a_1 a_2}} \text{ şi } r_0(X_1, X_2) = \frac{1}{2} \log \frac{a_1 a_2}{a_1 a_2 - h^2} \tag{7.2.4}$$

de unde rezultă relaţia (7.2.2).

Definiţia 7.2.1. Se numeşte coeficient de corelaţie informaţională, în sensul lui Linfoot, dintre variabilele aleatoare arbitrare X_1 şi X_2, expresia

$$L(X_1, X_2) = \sqrt{1 - e^{-2r_0(X_1, X_2)}} \tag{7.2.5}$$

Propoziţia 7.2.2. Pentru coeficientul de corelaţie informaţională $L(X_1, X_2)$ sînt adevărate următoarele afirmaţii

1) este definit pentru orice două variabile aleatoare X_1 şi X_2
2) este simetric în raport cu X_1 şi X_2, adică $L(X_1, X_2) = L(X_2, X_1)$
3) $0 \leqslant L(X_1, X_2) \leqslant 1$
4) $L(X_1, X_2) = 0$ dacă şi numai dacă X_1 şi X_2 sînt independente
5) $L(X_1, X_2) = 1$ dacă şi numai dacă X_1 şi X_2 sînt dependente funcţional
6) este invariant faţă de orice transformare reală biunivocă a variabilelor X_1 şi X_2, adică $L[\varphi_1(X_1), \varphi_2(X_2)] = L(X_1, X_2)$ pentru orice $\varphi_1, \varphi_2 : \mathbf{R} \to \mathbf{R}$ biunivoce.
7) dacă (X_1, X_2) este normal repartizat atunci

$$L^2(X_1, X_2) = \rho^2(X_1, X_2) \tag{7.2.6}$$

Scanned with OKEN Scanner

Demonstrație. Proprietățile 1) și 2) sînt evidente, 3) — 6) rezultă ca urmare a propoziției 7.1.1., iar cea de-a șaptea este o consecință a relației (7.1.9).

Se poate remarca imediat că, în raport cu postulatele lui Rényi, coeficientul de corelație informațională $L(X_1, X_2)$ poate fi socotit ca o bună măsură a dependenței, toate cele șapte axiome ale unui astfel de concept fiind satisfăcute.

7.3. J-divergența lui S. Kullback (1959)

Fie $\{\Omega, \mathcal{K}, \mu_i\}$, $i = 1, 2$ un cîmp de probabilitate, elementele lui Ω nefiind precizate, ele putînd fi unidimensionale sau multidimensionale, continue sau discrete, calitative sau cantitative. Vom presupune că μ_1 și μ_2 sînt echivalente ($\mu_1 \equiv \mu_2$) ceea ce înseamnă că ele sînt reciproc absolut continue (ceea ce înseamnă că $\mu_1(A) = 0 \Leftrightarrow \mu_2(A) = 0$, $A \in \mathcal{K}$ și scriem $\mu_1 \ll \mu_2$ și $\mu_2 \ll \mu_1$). Fie acum λ o altă măsură definită pe $\{\Omega, \mathcal{K}\}$ astfel că $\lambda \equiv \mu_1$ și $\lambda \equiv \mu_2$. Conform binecunoscutei teoreme a lui Radon-Nicodym există densitățile de probabilitate $\varphi_i(\omega)$, $i = 1, 2$, unice aproape peste tot în raport cu măsura λ, $0 < \varphi_i(\omega) < \infty$, astfel încît

$$\mu_i(E) = \int_E \varphi_i(\omega)\, d\lambda(\omega), \quad i = 1, 2 \quad E \in \mathcal{K} \tag{7.3.1}$$

Fie de asemenea Z o mărime aleatoare și H_1, H_2 două ipoteze asupra acesteia și anume

$$H_1 : P(Z \in \mathcal{K}) = \mu_1 \quad \text{și} \quad H_2 : P(Z \in \mathcal{K}) = \mu_2 \tag{7.3.2}$$

Conform formulei lui Bayes avem, pentru $i = 1, 2$

$$P(H_i/Z = \omega) = \frac{P(H_i)\, P(Z = \omega/H_i)}{P(H_1)P(Z = \omega/H_1) + P(H_2)P(Z = \omega/H_2)} =$$

$$= \frac{P(H_i)\, \varphi_i(\omega)}{P(H_1)\, \varphi_1(\omega) + P(H_2)\, \varphi_2(\omega)} \tag{7.3.3}$$

unde $P(H_i)$ înseamnă probabilitatea ca ipoteza H_i să fie adevărată apriori observației $Z = \omega$, iar $P(H_i/Z = \omega)$ probabilitatea ca H_i să fie adevărată aposteriori acestei observații. Din (7.3.3) deducem că

$$\log \frac{\varphi_1(\omega)}{\varphi_2(\omega)} = \log \frac{P(H_1/Z = \omega)}{P(H_2/Z = \omega)} - \log \frac{P(H_1)}{P(H_2)} \tag{7.3.4}$$

Scanned with OKEN Scanner

relație care exprimă cantitatea de informație obținută ca rezultat al observației $Z = \omega$ pentru a putea face o diferențiere între ipotezele H_1 și H_2 în favoarea ipotezei H_1, sau altfel spus, care exprimă ponderea sau greutatea argumentului introducerii observației $Z = \omega$ în favoarea ipotezei H_1 (S. Kullback (1959)).

Definiția 7.3.1. În sensul lui S. Kullback, mărimea

$$I[1:2;\ E] = \frac{1}{\mu_1(E)} \int_E \log \frac{\varphi_1(\omega)}{\varphi_2(\omega)}\, d\mu_1(\omega) =$$

$$= \begin{cases} \dfrac{1}{\mu_1(E)} \displaystyle\int_E \varphi_1(\omega) \log \frac{\varphi_1(\omega)}{\varphi_2(\omega)}\, d\lambda(\omega), & \mu_1(E) > 0 \\ 0 & \mu_1(E) = 0 \end{cases} \tag{7.3.5}$$

se numește cantitate medie de informație corespunzătoare diferențierii dintre ipotezele H_1 și H_2 în favoarea ipotezei H_1, în raport cu măsura μ_1.

Dacă $E = \Omega$ vom nota $I[1:2;\ \Omega] = I[1:2]$. Analog relației (7.3.5) se poate defini și mărimea $I[2:1]$. Cantitatea $I[1:2]$ mai este denumită și cantitatea medie de informație pe care o furnizează observațiile favorabile lui H_1, sau încă informația lui μ_1 în raport cu μ_2. Analog se interpretează și $I[2:1]$. Deducem imediat

Propoziția 7.3.1. Pentru $I[1:2]$ sînt adevărate următoarele

1) $I[1:2]$ este definită pentru orice μ_1 și μ_2 echivalente
2) $I[1:2] \neq I[2:1]$
3) $0 \leqslant I[1:2] \leqslant \infty$
4) $I[1:2] = 0$ dacă și numai dacă $\varphi_1(\omega) = \varphi_2(\omega)[\lambda]$
5) dacă $Z = (X_1, X_2)$ este un vector aleator ale cărui componente X_1 și X_2 sînt variabile aleatoare independente, atunci

$$I[1:2;\ X_1, X_2] = I[1:2;\ X_1] + I[1:2;\ X_2] \tag{7.3.6}$$

6) dacă $Z = (X_1, X_2)$ este un vector aleator arbitrar atunci

$$I[1:2;\ X_1, X_2] = I[1:2;\ X_i] + I[1:2;\ X_j/X_i]$$
$$1 \leqslant i \neq j \leqslant 2 \tag{7.3.7}$$

Scanned with OKEN Scanner

O proprietate remarcabilă a lui $I\,[1:2]$ este 4) care spune că $I[1:2]=0$ dacă cele două densităţi coincid aproape peste tot în raport cu măsura λ şi reciproc. Acest lucru înseamnă că între cele două repartiţii ale mărimii Z, sau între cele două ipoteze asupra acesteia, nu există nici o deosebire, deci nici o diferenţiere, ele coincizînd. Ca urmare a concavităţii funcţiei logaritmice se poate deduce că pentru orice $E\in\mathfrak{K}$ cu $\lambda(E)>0$ (Kullback (1959), Guiaşu (1977))

$$\int_E \varphi_1(\omega)\log\frac{\varphi_1(\omega)}{\varphi_2(\omega)}\,\mathrm{d}\lambda(\omega)\geqslant\left[\int_E\varphi_1(\omega)\mathrm{d}\lambda(\omega)\right]\left[\log\frac{\int_E\varphi_1(\omega)\mathrm{d}\lambda(\omega)}{\int_E\varphi_2(\omega)\mathrm{d}\lambda(\omega)}\right]=$$

$$=\mu_1(E)\log\frac{\mu_1(E)}{\mu_2(E)}\qquad(7.3.8)$$

cu egalitate dacă şi numai dacă $\dfrac{\varphi_1(\omega)}{\varphi_2(\omega)}=\dfrac{\mu_1(E)}{\mu_2(E)}\ [\lambda]$.

Ca urmare a acestui fapt se poate deduce de asemenea că pentru orice desfacere a lui Ω, $\{E_i\}_{i=1}^{\infty}$, $E_i\cap E_j=\emptyset$, $i\neq j$, $\bigcup_{i=1}^{\infty}E_i=\Omega$, are loc relaţia

$$I[1:2]\geqslant\sum_{i=1}^{\infty}\mu_1(E_i)\log\frac{\mu_1(E_i)}{\mu_2(E_i)}\qquad(7.3.9)$$

cu egalitate dacă şi numai dacă $\dfrac{\varphi_1(\omega)}{\varphi_2(\omega)}=\dfrac{\mu_1(E_i)}{\mu_2(E_i)}\ [\lambda]$, $\omega\in E_i$; $i\in\mathbf{N}$.

Dacă membrul stîng al relaţiei (7,3.8) reprezintă informaţia medie de diferenţiere pe care o conţin elementele lui E, iar membrul drept reprezintă informaţia de diferenţiere pe care o conţine însăşi E, atunci relaţia (7.3.8) arată că informaţia de diferenţiere adusă de părţi este mai mare decît cea furnizată de întregul E, egalitatea avînd loc numai cînd cele două ipoteze coincid şi nu mai avem ce diferenţia.

Definiţia 7.3.2. În sensul lui S. Kullback, mărimea

$$J(1,2)=I[1:2]+I[2:1]\qquad(7.3.10)$$

Scanned with OKEN Scanner

se numește măsură a divergenței dintre ipotezele H_1 și H_2, sau dintre repartițiile μ_1 și μ_2 sau încă J-divergență.

Ținînd seama de relația (7.3.5) pentru $E = \Omega$, putem scrie că

$$J(1,2) = \int_{\Omega} [\varphi_1(\omega) - \varphi_2(\omega)] \log \frac{\varphi_1(\omega)}{\varphi_2(\omega)} \, d\lambda(\omega). \qquad (7.3.11)$$

Ca urmare a celor de mai sus deducem

Propoziția 7.3.2. Pentru $J(1,2)$ sînt adevărate următoarele :

1) este definită pentru orice două măsuri echivalente μ_1 și μ_2

2) $J(1,2) = J(2,1)$

3) $0 \leqslant J(1,2) \leqslant \infty$

4) $J(1,2) = 0$ dacă și numai dacă $\varphi_1(\omega) = \varphi_2(\omega)$ $[\lambda]$

5) dacă $Z = (X_1, X_2)$ este un vector aleator, atunci

5.1) $J(1,2\,;\, X_1, X_2) = J(1,2\,;\, X_1) + J(1,2\,;\, X_2)$ dacă X_1 și X_2 sînt independente

5.2) $J(1,2\,;\, X_1, X_2) = J(1,2\,;\, X_i) + J(1,2\,;\, X_j/X_i)$, $1 \leqslant i \neq j \leqslant 2$, dacă X_1 și X_2 sînt dependente

6) pentru orice $E \in \mathfrak{K}$ cu $\lambda(E) > 0$

$$J(1,2\,;\, E) \geqslant [\mu_1(E) - \mu_2(E)] \log \frac{\mu_1(E)}{\mu_2(E)}, \qquad (7.3.12)$$

cu egalitate dacă și numai dacă $\dfrac{\varphi_1(\omega)}{\varphi_2(\omega)} = \dfrac{\mu_1(E)}{\mu_2(E)}$ $[\lambda]$

7) pentru orice desfacere a lui Ω, $\{E_i\}_{i=1}^{\infty}$, $E_i \cap E_j = \emptyset$, $i \neq j$, $\bigcup_{i=1}^{\infty} E_i = \Omega$

$$J(1,2) \geqslant \sum_{i=1} [\mu_1(E_i) - \mu_2(E_i)] \log \frac{\mu_1(E_i)}{\mu_2(E_i)} \qquad (7.3.13)$$

cu egalitate dacă și numai dacă $\dfrac{\varphi_1(\omega)}{\varphi_2(\omega)} = \dfrac{\mu_1(E_i)}{\mu_2(E_i)}$ $[\lambda]$, $\omega \in E_i$, $i \in \mathbf{N}$

Scanned with OKEN Scanner

Propoziția 7.3.3. Mărimile $I[1:2]$ și $I[2:1]$ satisfac întotdeauna relațiile

$$1)\quad \int_\Omega \log^2 \frac{\varphi_1(\omega)}{\varphi_2(\omega)}\, d\mu_2(\omega) + \int_\Omega \log \frac{\varphi_1(\omega)}{\varphi_2(\omega)}\, d\mu_2(\omega) \leqslant I[1:2] \leqslant$$

$$\leqslant \int_\Omega \left[\frac{\varphi_1(\omega) - \varphi_2(\omega)}{\varphi_2(\omega)}\right]^2 d\mu_2(\omega) \qquad (7.3.14)$$

$$2)\quad \int_\Omega \log^2 \frac{\varphi_2(\omega)}{\varphi_1(\omega)}\, d\mu_1(\omega) + \int_\Omega \log \frac{\varphi_2(\omega)}{\varphi_1(\omega)}\, d\mu_1(\omega) \leqslant I[2:1] \leqslant$$

$$\leqslant \int_\Omega \left[\frac{\varphi_1(\omega) - \varphi_2(\omega)}{\varphi_1(\omega)}\right]^2 d\mu_1(\omega) \qquad (7.3.15)$$

cu egalitate peste tot dacă și numai dacă $\varphi_1(\omega) = \varphi_2(\omega)$ $[\lambda]$.

Demonstrația este imediată dacă ținem cont că $\log x \leqslant x - 1$ sau că $1 + \log x \leqslant x$, $x > 0$. Ca urmare a propoziției 7.3.3. putem scrie că

$$\frac{1}{2}\int_\Omega [\varphi_1(\omega) + \varphi_2(\omega)]^2 \log^2 \frac{\varphi_1(\omega)}{\varphi_2(\omega)}\, d\lambda(\omega) \leqslant J(1,2) \leqslant$$

$$\leqslant \int_\Omega \frac{[\varphi_1(\omega) - \varphi_2(\omega)]^2[\varphi_1(\omega) + \varphi_2(\omega)]}{\varphi_1(\omega)\varphi_2(\omega)}\, d\lambda(\omega) \qquad (7.3.16)$$

cu egalitate dacă și numai dacă $\varphi_1(\omega) = \varphi_2(\omega)$ $[\lambda]$.

Deducem de asemenea imediat că

$$\int_\Omega \log^2 \frac{\varphi_i(\omega)}{\varphi_j(\omega)}\, d\mu_j(\omega) \leqslant J(1,2) \leqslant \int_\Omega \left[\frac{\varphi_i(\omega) - \varphi_j(\omega)}{\varphi_j(\omega)}\right]^2 d\mu_j(\omega),$$

$$1 \leqslant i \neq j \leqslant 2 \qquad (7.3.17)$$

cu egalitate peste tot dacă $\varphi_1(\omega) = \varphi_2(\omega)$ $[\lambda]$.

Să presupunem acum că $\Omega = \mathbf{R}^2$, $Z = (X_1, X_2)$, $\omega = (x_1, x_2)$, iar H_1 și H_2 două ipoteze asupra lui Z astfel că $H_1 : X_1$ și X_2 sînt dependente avînd repartiția comună caracterizată de măsura P_{X_1,X_2} definită de relația (6.3.3). $H_2 : X_1$ și X_2 sînt independente avînd repartiția comună caracterizată de măsura $P_{x_1 * x_2}$ definită de relația (6.3.5).

Scanned with OKEN Scanner

Prin urmare pentru orice mulțimi boreliene $C \subseteq \mathbf{R}^2$, $A \subseteq \mathbf{R}$, $B \subseteq \mathbf{R}$ avem $\mu_1(C) = P_{X_1,X_2}(C)$ și $\mu_2(A * B) = P_{X_1 * X_2}(A * B)$. În acest caz măsura λ va fi măsura Lebesgue. Ca urmare densitățile $\varphi_1(\omega)$ și $\varphi_2(\omega)$ sînt respectiv

$$\varphi_1(\omega) = \frac{d\mu_1(\omega)}{d\lambda(\omega)} = \frac{d^2 P_{X_1,X_2}(x_1, x_2)}{dx_1 dx_2} = f(x_1, x_2) \qquad (7.3.18)$$

$$\varphi_2(\omega) = \frac{d\mu_2(\omega)}{d\lambda(\omega)} = \frac{dP_{X_1}(x_1)}{dx_1} \cdot \frac{dP_{X_2}(x_2)}{dx_2} = g(x_1)h(x_2). \qquad (7.3.19)$$

În baza acestor considerații deducem cu ușurință prin calcul direct (Purcaru (1979))

Propoziția 7.3.4. Dacă (X_1, X_2) este repartizat normal $N(m_1, m_2, \sigma_1, \sigma_2, \rho)$, atunci

1) $$\int_\Omega \log^2 \frac{\varphi_1(\omega)}{\varphi_2(\omega)} d\mu_2(\omega) = \frac{1}{4} \log^2(1 - \rho^2) + \frac{\rho^2}{1 - \rho^2} \log(1 - \rho^2) + \frac{2\rho^2 + \rho^4}{(1 - \rho^2)^2} \qquad (7.3.20)$$

2) $$\int_\Omega \left[\frac{\varphi_1(\omega) - \varphi_2(\omega)}{\varphi_2(\omega)} \right]^2 d\mu_2(\omega) = \frac{\rho^2}{1 - \rho^2} \qquad (7.3.21)$$

3) $$I[1:2] = -\frac{1}{2} \log(1 - \rho^2) \qquad (7.3.22)$$

4) $$I[2:1] = \frac{1}{2} \log(1 - \rho^2) + \frac{\rho^2}{1 - \rho^2} \qquad (7.3.23)$$

5) $$J(1, 2) = C^2(X_1, X_2) = \frac{\rho^2}{1 - \rho^2} = e^{2I[1:2]} - 1 \qquad (7.3.24)$$

unde $C^2(X_1, X_2)$ este media pătratului de contingență (6.3.9).

6) $$\frac{1}{2}\rho^2 \leqslant I[1:2] \leqslant \frac{\rho^2}{2(1 - \rho^2)} \leqslant \frac{1}{2} J(1,2) \leqslant \frac{1 + \rho^2}{2} J(1, 2) \qquad (7.3.25)$$

cu egalitate peste tot dacă și numai dacă $\varphi_1(\omega) = \varphi_2(\omega)$ $[\lambda]$, adică dacă și numai dacă X_1 și X_2 sînt independente.

Tot prin calcul direct putem deduce (Purcaru (1979))

Scanned with OKEN Scanner

Propoziția 7.3.5. Dacă $\Omega = \mathrm{R}$, $Z = X_1$ și H_1, H_2 două ipoteze asupra repartiției variabilei aleatoare X_1 astfel încît

$$H_i : X_1 \text{ este normală } N(m_i, \sigma_i),\ i = 1, 2 \qquad (7.3.26)$$

atunci au loc întotdeauna relațiile

$$I[1:2]\log = \frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^4 + \sigma_1^2(m_1 - m_2)^2 - \sigma_1^2\sigma_2^2}{2\sigma_1^2\sigma_2^2} \qquad (7.3.27)$$

$$I[2:1] = \log\frac{\sigma_1}{\sigma_2} + \frac{\sigma_2^4 + \sigma_2^2(m_1 - m_2)^2 - \sigma_1^2\sigma_2^2}{2\sigma_1^2\sigma_2^2} \qquad (7.3.28)$$

$$J(1, 2) = \frac{(\sigma_1^2 - \sigma_2^2)^2 + (\sigma_1^2 + \sigma_2^2)(m_1 - m_2)^2}{2\sigma_1^2\sigma_2^2}. \qquad (7.3.29)$$

Demonstrația nu este dificilă deși unele calcule sînt mai lungi. Într-adevăr, să observăm că $\log\frac{\varphi_1(\omega)}{\varphi_2(\omega)} = \log\frac{\sigma_2}{\sigma_1} - \frac{1}{2}\left[\left(\frac{x - m_1}{\sigma_1}\right)^2 - \left(\frac{x - m_2}{\sigma_2}\right)^2\right]$ și ca urmare putem scrie că

$$I[1:2] = \int_{\mathrm{R}} \varphi_1 \log\frac{\varphi_1}{\varphi_2}\,dx = \log\frac{\sigma_2}{\sigma_1} - \frac{1}{2}\int_{\mathrm{R}}\left[\left(\frac{x - m_1}{\sigma_1}\right)^2 - \left(\frac{x - m_2}{\sigma_2}\right)^2\right]\varphi_1(x)dx$$

de unde prin efectuarea calculelor rezultă relația (7.3.27). Analog se verifică și celelalte. Ca o consecință imediată a propoziției 7.3.5 putem formula

Propoziția 7.3.6. În condițiile propoziției 7.3.5

1) dacă $\sigma_1^2 = \sigma_2^2$ și $m_1 = m_2$ simultan atunci și numai atunci

$$I[1:2] = I[2:1] = J(1, 2) = 0 \qquad (7.3.30)$$

2) dacă $\sigma_1^2 = \sigma_2^2 = \sigma^2$ și $m_1 \neq m_2$, atunci

$$I[1:2] = I[2:1] = \frac{1}{2}J(1, 2) = \frac{(m_1 - m_2)^2}{2\sigma^2} \qquad (7.3.31)$$

3) dacă $\sigma_1^2 \neq \sigma_2^2$ și $m_1 = m_2$, atunci

$$I[1:2] = \log\frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 - \sigma_2^2}{2\sigma_2^2};$$

Scanned with OKEN Scanner

$$I[2:1] = \log \frac{\sigma_1}{\sigma_2} + \frac{\sigma_2^2 - \sigma_1^2}{2\sigma_1^2};$$

$$J(1, 2) = \frac{(\sigma_1^2 - \sigma_2^2)^2}{2\sigma_1^2\sigma_2^2} \tag{7.3.32}$$

4) dacă $m_1 = m$, $\sigma_1^2 = \sigma^2$ și $m_2 = 0$, $\sigma_2 = 1$, atunci

$$I([1:2] = \log \frac{1}{\sigma} + \frac{\sigma^2 + m^2 - 1}{2};$$

$$I[2:1] = \log \sigma + \frac{m^2 + 1 - \sigma^2}{2\sigma^2};$$

$$J(1, 2) = \frac{(1 - \sigma^2)^2 + m^2(1 + \sigma^2)}{2\sigma^2}. \tag{7.3.33}$$

Ca o generalizare imediată a celor de mai sus putem da (Purcaru (1979))

Propoziția 7.3.7. Dacă $\Omega = \mathbf{R}^n$, $Z = X = (X_1, \ldots, X_n)$, $\omega = (x_1, \ldots, x_n)$ și H_1, H_2 două ipoteze asupra lui X astfel că H_1 : vectorul aleator X are repartiția

$$\varphi_1(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2} \prod_{k=1}^{n} \sigma_k} \exp\left\{-\frac{1}{2} \sum_{k=1}^{n} \left(\frac{x_k - m_k}{\sigma_k}\right)^2\right\} \tag{7.3.34}$$

H_2 : vectorul aleator X are repartiția

$$\varphi_2(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2} \prod_{k=1}^{n} \bar{\sigma}_k} \exp\left\{-\frac{1}{2} \sum_{k=1}^{n} \left(\frac{x_k - \bar{m}_k}{\bar{\sigma}_k}\right)^2\right\} \tag{7.3.35}$$

atunci au loc egalitățile

$$1)\ I[1:2] = \sum_{k=1}^{n} \log \frac{\bar{\sigma}_k}{\sigma_k} + \frac{1}{2} \sum_{k=1}^{n} \frac{\sigma_k^2 + (m_k - \bar{m}_k)^2}{\bar{\sigma}_k^2} - \frac{n}{2} \tag{7.3.36}$$

$$2)\ I[2:1] = \sum_{k=1}^{n} \log \frac{\sigma_k}{\bar{\sigma}_k} + \frac{1}{2} \sum_{k=1}^{n} \frac{\bar{\sigma}_k^2 + (m_k - \bar{m}_k)^2}{\sigma_k^2} - \frac{n}{2} \tag{7.3.37}$$

Scanned with OKEN Scanner

$$3)\ J(1,\ 2) = \frac{1}{2}\sum_{k=1}^{n} \frac{(\sigma_k^2 - \bar{\sigma}_k^2)^2 + (m_k - \bar{m}_k)^2(\sigma_k^2 + \bar{\sigma}_k^2)}{\sigma_k^2 \bar{\sigma}_k^2} \tag{7.3.38}$$

Demonstrația se obține prin calcul direct și se observă că dacă $n = 1$ relațiile (7.3.27) — (7.3.29) și (7.3.36) — (7.3.38) coincid. Particularizînd parametrii m, $\bar{m}$, σ și $\bar{\sigma}$ în propoziția 7.3.7. putem generaliza rezultatele din propoziția 7.3.6.

Propoziția 7.3.8. În condițiile propoziției 7.3.7, pentru orice $1 \leqslant k \leqslant n$

1) dacă $\sigma_k = \bar{\sigma}_k$ și $m_k = \bar{m}_k$ simultan, atunci și numai atunci

$$I[1:2] = I[2:1] = J(1,\ 2) = 0 \tag{7.3.39}$$

2) dacă $\sigma_k = \bar{\sigma}_k$ și $m_k \neq \bar{m}_k$ atunci

$$I[1:2] = I[2:1] = \frac{1}{2} J(1,\ 2) = \frac{1}{2}\sum_{k=1}^{n} \frac{(m_k - \bar{m}_k)^2}{\sigma_k^2} \tag{7.3.40}$$

3) dacă $\sigma_k \neq \bar{\sigma}_k$ și $m_k = \bar{m}_k$, atunci

$$I[1:2] = \sum_{k=1}^{n} \log \frac{\bar{\sigma}_k}{\sigma_k} + \frac{1}{2}\sum_{k=1}^{n} \frac{\sigma_k^2 - \bar{\sigma}_k^2}{\bar{\sigma}_k^2};$$

$$I[2:1] = \sum_{k=1}^{n} \log \frac{\sigma_k}{\bar{\sigma}_k} + \frac{1}{2}\sum_{k=1}^{n} \frac{\bar{\sigma}_k^2 - \sigma_k^2}{\sigma_k^2};$$

$$J(1,\ 2) = \sum_{k=1}^{n} \frac{(\sigma_k^2 - \bar{\sigma}_k^2)^2}{2\sigma_k^2 \bar{\sigma}_k^2} \tag{7.3.41}$$

4) dacă $\bar{m}_k = 0$, $\bar{\sigma}_k = 1$, atunci

$$\left.\begin{aligned} I[1:2] &= -\sum_{k=1}^{n} \log \sigma_k + \frac{1}{2}\sum_{k=1}^{n} (\sigma_k^2 + m_k^2 - 1); \\ I[2:1] &= \sum_{k=1}^{n} \log \sigma_k + \frac{1}{2}\sum_{k=1}^{n} \frac{1 + m_k^2 - \sigma_k^2}{\sigma_k^2}; \\ J(1,\ 2) &= \frac{1}{2}\sum_{k=1}^{n} \frac{(\sigma_k^2 - 1)^2 + m_k^2(\sigma_k^2 + 1)}{\sigma_k^2} \end{aligned}\right\} \tag{7.3.42}$$

5) dacă $\bar{m}_k = 0$, $\bar{\sigma}_k = 1$, $m_k = m$, $\sigma_k = \sigma$, atunci

$$I[1:2] = -n \log \sigma + \frac{n}{2}(\sigma^2 + m^2 - 1);$$

$$I[2:1] = n \log \sigma + \frac{n}{2}\left(\frac{1 + m^2 - \sigma^2}{\sigma^2}\right)$$

Scanned with OKEN Scanner

$$J(1,2) = \frac{n}{2}\frac{(\sigma^2 - 1)^2 + m^2(\sigma^2 + 1)}{\sigma^2}. \qquad (7.3.43)$$

Deoarece valoarea maximă a lui $J(1,2)$ poate fi oricît de mare, putem introduce o valoare normată a acesteia (Purcaru 1979).

Definiția 7.3.3. Se numește coeficient de divergență raportul

$$\bar{J}(1, 2) = \frac{J(1, 2)}{1 + J(1, 2)} \qquad (7.3.44)$$

Ca urmare a proprietăților J-divergenței prezentate mai sus deducem (Purcaru (1979))

Propoziția 7.3.9. Coeficientul de divergență satisface următoarele

1) este definit pentru orice două măsuri echivalente μ_1 și μ_2
2) este simetric în raport cu μ_1 și μ_2, adică $\bar{J}(1, 2) = \bar{J}(2, 1)$
3) $0 \leqslant \bar{J}(1, 2) \leqslant 1$
4) $\bar{J}(1, 2) = 0$ dacă și numai dacă $\varphi_1(\omega) = \varphi_2(\omega)$ $[\lambda]$
5) $\bar{J}(1, 2) \to 1$ dacă și numai dacă $J(1, 2) \to \infty$
6) este invariant față de orice transformare reală biunivocă ξ a lui Z, adică

$$\bar{J}(1, 2; \xi(Z)) = J(1, 2; Z) \qquad (7.3.45)$$

7) dacă $\Omega = \mathbf{R}^2$ și $Z = (X_1, X_2)$ este repartizat normal bidimensional, atunci

$$\bar{J}(1, 2; X_1, X_2) = \rho^2(X_1, X_2) \qquad (7.3.46)$$

Remarcăm că pentru coeficientul de divergență $\bar{J}(1, 2; X_1, X_2)$ sînt verificate într-un caz mai general, postulatele lui Rényi. Dacă acesta este considerat în ipotezele premergătoare relațiilor (7.3.18) și (7.3.19), atunci proprietățile lui din propoziția (7.3.9) sînt chiar postulatele lui Rényi. Acest fapt remarcabil ne îndreptățește să considerăm J-divergența și coeficientul de divergență $\bar{J}(1, 2)$ ca măsuri ale dependenței într-un sens mai larg, acela al dependenței dintre două distribuții probabilistice.

Scanned with OKEN Scanner

Să presupunem acum că $\Omega = \mathbf{R}^n$, $Z = X = (X_1, \ldots, X_n)$, $\omega = (x_1, x_2, \ldots, x_n)$, iar H_1 și H_2 sînt două ipoteze asupra lui X astfel că

H_1 : X are componentele dependente două cîte două cu densitatea de repartiție $\varphi_1^{kj}(\omega)$, $1 \leqslant j, \ k \leqslant n$

H_2 : X are componentele independente două cîte două cu densitatea de repartiție $\varphi_2^{kj}(\omega)$, $1 \leqslant j, k \leqslant n$

Ca urmare, prin analogie cu matricea covariantă și cu cea de corelație $\Sigma(X)$ și $\mathcal{R}(X)$ din definiția 6.1.4, putem introduce matricele

$$\mathcal{J}(1, 2) = (J_{jk}(1, 2)) \quad \text{și} \quad \overline{\mathcal{J}}(1, 2) = (J_{jk}(1, 2)) \tag{7.3.47}$$

ale căror elemente sînt J-divergența și respectiv coeficientul de divergență corespunzînd perechilor de variabile aleatoare (X_j, X_k) în ipotezele menționate. Pentru fiecare pereche în parte H_1 și H_2 sînt exprimate de relațiile (7.3.18) și (7.3.19). Ca urmare, propozițiile 7.3.2 și 7.3.9 pot fi extinse pentru matricele (7.3.47). Se poate remarca imediat că dacă X este repartizat normal atunci

$$\overline{\mathcal{J}}(1, 2) = \mathcal{R}(X) \tag{7.3.48}$$

Matricea $\mathcal{J}(1, 2)$ se numește matrice de divergență, iar $\overline{\mathcal{J}}(1, 2)$ matricea coeficienților de divergență.

O extensie a mărinilor P_{X_1,X_2} și $P_{X_1 * X_2}$ definite de relațiile (6.3.3) și (6.3.5) la cazul unui vector aleator n-dimensional și în consecință formularea ipotezelor H_1 și H_2 astfel încît să se extindă relațiile (7.3.18) și (7.3.19) ne conduc la a scrie că

$$J(1, 2) = \int_{\mathbf{R}^n} \left[f(x_1, \ldots, x_n) - \prod_{k=1}^{n} f_k(x_k) \right] \times$$

$$\times \log \frac{f(x_1, \ldots, x_n)}{\prod_{k=1}^{n} f_k(x_k)} \, \mathrm{d}x_1 \ldots \mathrm{d}x_n \tag{7.3.49}$$

Ținînd cont de proprietățile J-divergenței, relația (7.3.49) definește o măsură globală a dependenței dintre

Scanned with OKEN Scanner

componentele vectorului aleator X, ilustrînd încă o dată faptul că J-divergența și coeficientul de divergență pot fi socotite și utilizate în consecință ca măsuri ale dependenței probabiliste și chiar ca bune măsuri ale dependenței în sensul lui Rényi. Rezultate asemănătoare celor din propozițiile 7.3.5 — 7.3.6 pot fi obținute fără dificultate și pentru celelalte repartiții clasice. Ele apar de altfel ca simple exerciții ilustrative cînd asupra unei repartiții clasice facem două ipoteze asupra parametrului sau parametrilor acesteia.

7.4. Coeficientul de corelație infomațională al lui C. B. Bell (1962)

Preocupat de definirea unor măsuri ale dependenței care să satisfacă postulatele lui Rényi privind conceptul de „bună măsură a dependenței" și pornind de la coeficientul logaritmic de corelație (7.1.7), C.B.Bell a definit două măsuri asemănătoare și a propus unele reformulări ale postulatelor lui Rényi.

Fie X_1 și X_2 două variabile aleatoare și $H(X_1)$, $H(X_2)$ și $H(X_1, X_2)$ entropiile Shannon corespunzătoare.

Definiția 7.4.1. Se numesc coeficienți de corelație informațională a variabilelor aleatoare X_1 și X_2, în sensul C.B.Bell, rapoartele

$$B_1(X_1, X_2) = \frac{H(X_1) + H(X_2) - H(X_1, X_2)}{\min [H(X_1), H(X_2)]} \qquad (7.4.1)$$

$$B_2(X_1, X_2) = \frac{H(X_1) + H(X_2) - H(X_1, X_2)}{\max [H(X_1), H(X_2)]} \qquad (7.4.2)$$

Ținînd cont de proprietățile entropiei Shannon, deducem fără dificultate

Propoziția 7.4.1. Pentru $B_1(X_1, X_2)$ și $B_2(X_1, X_2)$ sînt adevărate următoarele

1) sînt definiți pentru orice vector aleator (X_1, X_2) cu componente neconstante
2) $B_i(X_1, X_2) = B_i(X_2, X_1)$, $i = 1, 2$
3) $0 \leqslant B_2(X_1, X_2) \leqslant B_1(X_1, X_2) \leqslant 1$
4) $B_i(X_1, X_2) = 0$ dacă și numai dacă X_1 și X_2 sînt independente

Scanned with OKEN Scanner

5) $B_i(X_1, X_2) = 1$ dacă și numai dacă cealaltă componentă este complet dependentă de componenta ce dă numitorul lui $B_i(X_1, X_2)$

6) sînt invarianți în raport cu orice transformare reală biunivocă a variabilelor X_1 și X_2, $\varphi = (\varphi_1, \varphi_2)$, adică

$$B_i[\varphi_1(X_1), \varphi_2(X_2)] = B_i(X_1, X_2).$$

Să remarcăm că putem alege oricînd X_1 și X_2 astfel încît

$$B_i(X_1, X_2) = \frac{H(X_1) + H(X_2) - H(X_1, X_2)}{H(X_i)} =$$

$$= 1 - \frac{H(X_i/X_j)}{H(X_i)}, \quad 1 \leqslant 1 \neq j \leqslant 2 \qquad (7.4.3)$$

avînd spre exemplu $H(X_1) \leqslant H(X_2)$.

Se observă că $B_i(X_1, X_2)$ nu satisfac cel de-al șaptelea postulat al lui Rényi, iar cel de-al cincilea este satisfăcut acum cu „dacă și numai dacă". Din aceste motive C. B. Bell a propus ca aceste două postulate să fie reformulate astfel:

(R.V′) $\delta(X_1, X_2) = 1$ dacă și numai dacă între X_1 și X_2 există o dependență strictă adică există funcțiile bijective f_1 și f_2 astfel că fie $X_1 = f_2(X_2)$, fie $X_2 = f_1(X_1)$

(R.VII′) $\delta(X_1, X_2)$ este o funcție strict monotonă de $|\rho(X_1, X_2)|$ dacă X_1 și X_2 sînt normal repartizate.

Se poate verifica imediat că dacă (X_1, X_2) este repartizat normal atunci

$$B_i(X_1, X_2) = -\frac{\log(1 - \rho^2)}{\log 2\pi e \sigma^2}, \quad i = 1, 2 \qquad (7.4.4)$$

ceea ce justifică (R.VII′). Modificarea postulatului al șaptelea nu afectează măsurile celelalte care satisfac (R.VII) căci relația $\delta(X_1, X_2) = |\rho(X_1, X_2)|$, după cum $\rho > 0$ sau $\rho < 0$, exprimă un caz particular de monotonie strictă a lui $\delta(X_1, X_2)$ ca funcție de ρ.

Vom spune că o măsură a dependenței care satisface primele șase postulate ale lui Rényi și pe cel de-al șaptelea în versiunea lui Bell, se numește „bună măsură" a dependenței în sensul lui Rényi-Bell.

Scanned with OKEN Scanner

7.5. Energia și corelația informațională a lui O. Onicescu (1966)

Fie X o variabilă aleatoare discretă, ale cărei valori reprezintă stările unui sistem oarecare, avînd probabilitățile

$$p_k \geqslant 0, \quad \sum_{k=1}^{n} p_k = 1.$$

Definiția 7.5.1. Se numește energie informațională, în sensul lui O. Onicescu, corespunzătoare repartiției variabilei aleatoare discrete X, expresia

$$E^2(X) = \sum_{k=1}^{n} p_k^2. \tag{7.5.1}$$

Energia informațională $E^2(X)$ apare ca o valoare globală atașată unei repartiții probabiliste, de aceeași natură ca și entropia lui Shannon (O. Onicescu (1973)). Se poate demonstra fără dificultate (O. Onicescu (1966, 1972, 1973))

Propoziția 7.5.1. Energia informațională $E^2(X)$ satisface următoarele :

1) este definită pentru orice variabilă aleatoare discretă

2) $\frac{1}{n} \leqslant E^2(X) \leqslant 1$

3) $E^2(X) = \frac{1}{n}$ dacă și numai dacă X este uniform repartizată

4) $E^2(X) = 1$ dacă și numai dacă avem o concentrare absolută informațională, adică $p_k = 1$, $p_j = 0$, $j \neq k$, $1 \leqslant j \leqslant n$

5) energia informațională scade cînd nedeterminarea sau uniformitatea crește

6) dacă variabila $X' = (x_1', \ldots, x_m')$ provine din X, $m \leqslant n$, atunci $E^2(X') \leqslant E(X)$

7) este invariantă față de orice transformare biunivocă a lui X

8) dacă X_1 și X_2 sînt două variabile aleatoare discrete independete atunci $E^2(X_1X_2) = E^2(X_1)E^2(X_2)$, iar dacă sînt dependente atunci $E^2(X_1X_2) = E^2(X_j)E^2(X_k|X_j)$ $1 \leqslant j \neq k \leqslant 2$

Definiția 7.5.2. Se numește corelație informațională a variabilelor aleatoare X_1 și X_2, sau încă, a repartițiilor pro-

Scanned with OKEN Scanner

babilistice $p \geqslant 0$ și $q_k \geqslant 0$, $\sum_{k=1}^{n} p_k = \sum_{k=1}^{n} q_k = 1$, în sensul lui O. Onicescu, mărimea

$$O^2[X_1, X_2] = \sum_{k=1}^{n} p_k q_k \tag{7.5.2}$$

și coeficient de corelație informațională al acestora, raportul

$$R^2[X_1, X_2) = \frac{O^2[X_1, X_2]}{\sqrt{E^2(X_1)E^2(X_2)}} \tag{7.5.3}$$

Variabilele aleatoare X_1 și X_2 pot fi două variabile arbitrare în ceea ce privește conținutul acestora sau pot reprezenta o aceeași caracteristică aleatoare dar pentru două colectivități sau populații statistice diferite. Poate fi vorba prin urmare de a studia gradul de asociere dintre cele două repartiții $p_k \geqslant 0$ și $q_k \geqslant 0$ ale aceleiași variabile aleatoare aflate în studiu. Pornind de la definiția expresiilor (7.5.2) și (7.5.3) se poate demonstra fără dificultate (Onicescu (1972, 1973))

Propoziția 7.5.2. Pentru mărimile $O^2[X_1, X_2]$ și $R^2[X_1, X_2]$ sînt adevărate următoarele afirmații

1) sînt definite pentru orice pereche de repartiții aleatoare $p_k \geqslant 0$ și $q_k \geqslant 0$, $\sum_{k=1}^{n} p_k = \sum_{k=1}^{n} q_k = 1$;

2) sînt simetrice în raport cu cele două repartiții probabilistice;

3) $\max(\min_k p_k, \min_k q_k) \leqslant O^2[X_1, X_2] \leqslant \min(\max_k p_k, \max_k q_k)$ și analog pentru $R^2[X_1, X_2]$ după împărțirea peste tot cu $E(X_1)E(X_2)$, cu egalitate peste tot dacă și numai dacă ambele repartiții sînt uniforme;

4) $0 \leqslant O^2[X_1, X_2] \leqslant 1$ și $0 \leqslant R^2[X_1, X_2] \leqslant 1$;

5) $O^2[X_1, X_2] = R^2[X_1, X_2] = 0$ dacă și numai dacă colectivitățile statistice cărora le corespund X_1 și X_2 se află într-o relație de străinătate în ceea ce privește starea k a acestora (valoarea x_1^k sau x_2^k fiind în fapt aceeași pentru ambele, dar diferind repartiția ei), adică cel puțin una din probabilitățile p_k sau q_k este nulă, $1 \leqslant k \leqslant n$;

6) $O^2[X_1, X_2] = 1$ dacă și numai dacă aceeași stare x^k este exclusivă pentru ambele populații, adică $p_k = q_k = 1$,

Scanned with OKEN Scanner

iar $R^2[X_1, X_2] = 1$ dacă și numai dacă cele două repartiții sînt identice.

Se poate remarca fără dificultate faptul că nespecificînd dimensiunea caracteristicii aleatoare în discuție, poate fi vorba de definirea mărimilor (7.5.1), (7.5.2) și (7.5.3) atît pentru variabile aleatoare unidimensionale cît și pentru cele multidimensionale. De asemenea aceste mărimi pot fi definite pentru repartiții continue ale căror densități de repartiție sînt funcții de pătrat integrabil (Onicescu (1972, 1973)).

Generalizînd definiția 7.5.1, putem introduce conceptul de energie informațională de ordin α, pe care l-am definit prin relația (3.5.55), concept strîns legat de α-entropia lui Rényi și alfa-entropia Behara-Nath după cum se poate vedea din propoziția 3.5.16. Așadar numim energie informațională Onicescu de ordin α, $\alpha > 0$, sau generalizată, expresia (I. Purcaru (1988))

$$E^{\alpha}(X) = \sum_{k=1}^{n} p_k^{\alpha} \tag{7.5.4}$$

observînd că dacă $\alpha = 2$, relațiile (7.5.1) și (7.5.4) coincid, iar pentru $\alpha = 1$ avem relația trivială a completitudinii repartiției $p_k \geqslant 0$, $\sum_{k=1}^{n} p_k = 1$. Deducem fără dificultate

Propoziția 7.5.3. Avem întotdeauna

1) $0 < E^{\alpha}(X) \leqslant n^{1-\alpha}$, dacă $0 < \alpha \leqslant 1$, cu egalitate dacă și numai dacă X este repartizată uniform;

2) $n^{1-\alpha} \leqslant E^{\alpha}(X) \leqslant 1$, dacă $\alpha > 1$, cu primă egalitate ca la 1) și cu a doua dacă și numai dacă avem o concentrare informațională absolută, adică $p_k = 1$ și $p_j = 0$, $j \neq k$, $1 \leqslant j \leqslant n$.

Se poate remarca faptul că pentru $E^{\alpha}(X)$ putem formula o propoziție analogă propoziției 7.5.1 Relația (7.5.4) poate fi dată mai general, chiar în cazul unei repartiții incomplete ($p_k \geqslant 0$, $\sum_{k=1}^{n} p_k \leqslant 1$). Analog generalizării (7.5.4) putem da

Definiția 7.5.3. Se numește corelație informațională Onicescu generalizată corespunzătoare repartițiilor $p_k \geqslant 0$, $q_k \geqslant 0$, $\sum_{k=1}^{n} p_k = \sum_{k=1}^{n} q_k = 1$, ale variabilelor X_1 și X_2, expresia

$$O^{\alpha}[X_1, X_2] = \sum_{k=1}^{n} p_k^{\frac{\alpha}{2}} q_k^{\frac{\alpha}{2}}, \quad \alpha > 0 \tag{7.5.5}$$

Scanned with OKEN Scanner

și coeficient de corelație informațională generalizată a acestora, raportul

$$R^{\alpha}[X_1, X_2] = \frac{O^{\alpha}[X_1, X_2]}{[E^{\alpha}(X_1)E^{\alpha}(X_2)]^{1/2}}, \quad \alpha > 0. \tag{7.5.6}$$

Pentru trei caracteristici, deci pentru trei repartiții $p_k \geqslant 0$, $q_k \geqslant 0$, $r_k \geqslant 0$, complete, putem scrie corelația informațională generalizată

$$O^{\alpha}[X_1, X_2, X_3] = \sum_{k=1}^{n} p_k^{\frac{\alpha}{3}} q_k^{\frac{\alpha}{3}} r_k^{\frac{\alpha}{3}} \tag{7.5.7}$$

și coeficientul de corelație informațională corespunzător ca fiind

$$R^{\alpha}[X_1, X_2, X_3] = \frac{O^{\alpha}[X_1, X_2, X_3]}{[E^{\alpha}(X_1)E^{\alpha}(X_2)E^{\alpha}(X_3)]^{1/3}} \tag{7.5.8}$$

relații care pentru $\alpha = 3$ coincid cu cele introduse de O. Onicescu (1973).

7.6. Corelație și coeficienți de corelație entropică (1969, 1973, 1976)

Ideea definirii și studierii măsurilor entropice ale dependenței, denumire pe care am adoptat-o spre a se deosebi de celelalte măsuri informaționale, a fost furnizată cu siguranță de proprietățile entropiei Shannon, cu precădere cea legată de independența entropică și probabilistică a două variabile aleatoare. Prima încercare în acest sens este coeficientul logaritmic de corelație al spaniolilor M. C. Camargo și I. Medina (1955) care i-a inspirat pe E. H. Linfoot (1957) C. B. Bell (1962) și cu siguranță pe S. Watanabe (1969), care a propus drept măsură a independenței sau a coeziunii dintre un anumit număr de mărimi aleatoare, diferența dintre suma entropiilor acestora și entropia ansamblului pe care acestea îl constituie. Pornind de la această idee s-au definit și studiat diverse măsuri entropice ale dependenței înglobate sub numele de corelație entropică.

7.6.1. Măsuri entropice ale dependenței globale

7.6.1.1. Dependența a două variabile aleatoare. Fie X_1 și X_2 două variabile aleatoare discrete și entropiile corespunzătoare acestora $H(X_1, X_2)$, $H(X_i)$ și $H(X_i/X_j)$

$1 \leqslant l \neq j \leqslant 2$ definite în paragraful 3.3. Să considerăm acum expresia

$$W[X;\ X_1,\ X_2] = H(X_1) + H(X_2) - H(X). \qquad (7.6.1)$$

Ținînd cont de relația (3.3.9) avem

$$W[X;\ X_1,\ X_2] = H(X_i) - H(X_i/X_j),\ 1 \leqslant i \neq j \leqslant 2 \qquad (7.6.2)$$

Ca urmare a propoziției 3.3.5 deducem fără dificultate

Propoziția 7.6.1. Următoarele afirmații asupra mărimii $W[X;\ X_1,\ X_2]$ sînt adevărate totdeauna

1) este definită pentru orice vector aleator $(X_1,\ X_2)$;

2) $W[X;\ X_1;\ X_2] = W[X;\ X_2,\ X_1]$;

3) $0 \leqslant W[X;\ X_1,\ X_2] < \infty$;

4) $W[X;\ X_1,\ X_2] = 0$ dacă și numai dacă X_1 și X_2 sînt independente;

5) $W[X;\ X_1,\ X_2] = H(X_i)$ dacă și numai dacă X_i c.d. X_j, $1 \leqslant i \neq j \leqslant 2$;

6) $W[X;\ X_1,\ X_2] = H(X_1) = H(X_2) = H(X_1,\ X_2)$ dacă și numai dacă X_1 r.c.d. X_2;

7) este invariantă față de orice transformare reală biunivocă a variabilelor X_1 și X_2.

Ca urmare a propoziției 3.3.6 putem enunța

Propoziția 7.6.2. Pentru orice $(X_1,\ X_2)$ avem

$$W[X;\ X_1X_2] \leqslant \min[H(X_1),\ H(X_2)] \leqslant$$

$$\leqslant \frac{1}{2}\,[H(X_1) + H(X_2)] \leqslant \max[H(X_1),\ H(X_2)] \leqslant$$

$$\leqslant H(X_1,\ X_2) \qquad (7.6.3)$$

cu egalitate peste tot dacă și numai dacă X_1 r.c.d. X_2.

Definiția 7.6.1. Mărimea $W[X;\ X_1,\ X_2]$ se numește corelație entropică globală a variabilelor aleatoare X_1 și X_2.

Să notăm cu $M_i[X_1,\ X_2]$, $1 \leqslant i \leqslant 4$, cei patru majoranți ai corelației entropice $W[X;\ X_1,\ X_2]$ dați de relația (7.6.3), în ordine descrescătoare.

Scanned with OKEN Scanner

Definiția 7.6.2. Se numesc coeficienți de corelație entropică ai variabilelor aleatoare X_1 și X_2 rapoartele

$$G_i[X; X_1, X_2] = \frac{W[X; X_1, X_2]}{M_i[X_1, X_2]}, \quad i \leqslant i \leqslant 4. \quad (7.6.4)$$

Ca o consecință imediată a propozițiilor 7.6.1 și 7.6.2 putem enunța

Propoziția 7.6.3. Pentru coeficienții de corelație entropică $G_i[X; X_1, X_2]$ sînt adevărate următoarele:

1) sînt definiți pentru orice vector aleator (X_1, X_2), nici una din componentele sale nefiind constantă cu probabilitatea unu;
2) $G_i[X; X_1, X_2] = G_i[C; X_2, X_1]$, $1 \leqslant i \leqslant 4$;
3) $0 \leqslant G_i[X; X_1, X_2] \leqslant G_{i+1}[X; X_1, X_2] \leqslant 1$, $1 \leqslant i \leqslant 3$
4) $G_i[X; X_1, X_2] = 0$ dacă și numai dacă X_1 și X_2 sînt independente
5) $G_i[X; X_1, X_2] = 1$, $1 \leqslant i \leqslant 4$, dacă și numai dacă X_1 r.c.d. X_2
6) sînt invarianți față de orice transformare reală biunivocă $\varphi_1, \varphi_2 : \mathbf{R} \to \mathbf{R}$ a variabilelor X_1 și X_2, adică,

$$G_i[\varphi(X); \varphi_1(X_1), \varphi_2(X_2)] = G_i[X; X_1, X_2], \quad 1 \leqslant i \leqslant 4.$$

Să remarcăm că pentru $i = 2$ și $i = 4$, $G_i[X; X_1, X_2] = B_i(X_1, X_2)$, adică tocmai coeficienții lui Bell (7.4.1) și (7.4.2), iar $G_3[X; X_1, X_2] = R_0(X_1, X_2)$, adică tocmai coeficientul logaritmic de corelație (7.1.7). Remarcăm de asemenea că, exceptînd cel de al șaptelea postulat al lui Rényi, toate celelalte șase postulate privind conceptul de bună măsură a dependenței sînt satisfăcute de către coeficienții de corelație entropică. Ideea și implicit coeficientul lui Linfoot (7.2.5) ne sugerează și posibilitatea soluționării unei astfel de probleme.

Următoarea proprietate a coeficienților $G_i[X; X_1, X_2]$, deși simplă, este remarcabilă.

Propoziția 7.6.4. Orice combinație liniară convexă

$$G[X; X_1, X_2; \alpha] = \sum_{i=1}^{4} \alpha_i G_i[X; X_1, X_2], \ \alpha_i \geqslant 0, \ \sum_{i=1}^{4} \alpha_i = 1 \quad (7.6.5)$$

a coeficienților de corelație entropică este de asemenea un coeficient de corelație entropică.

Scanned with OKEN Scanner

Demonstrație. Pentru a demonstra afirmația este suficient să arătăm că $G[X; X_1, X_2; \alpha]$ satisface proprietățile din propoziția 7.6.3 — Proprietățile 1) — 4) și 6) sînt evidente. Pentru cea de a cincea, remarcăm că dacă X_1 r.c.d. X_2 atunci toți $G_i[X; X_1, X_2] = 1$ și implicit $G[X; X_1, X_2; \alpha] = 1$. Reciproc, dacă $G[X; X_1, X_2; \alpha] = 1$ atunci

$$\sum_{i=1}^{4} \alpha_i G_i[X; X_1, X_2] = \sum_{i=1}^{4} \alpha_i \Rightarrow \sum_{i=1}^{4} \alpha_i (1 - G_i[X; X_1, X_2]) = 0.$$

Cum $0 \leqslant \alpha_i \leqslant 1$, rezultă că $G_i[X; X_1, X_2] = 1$, $1 \leqslant i \leqslant 4$, deci X_1 r.c.d X_2 și demonstrația este terminată. Am presupus de fapt $0 < \alpha_i < 1$ căci dacă spre exemplu $\alpha_1 = \alpha_2 = \alpha_3 = 0$ și $\alpha_4 = 1$ atunci $G[X; X_1, X_2; \alpha] = G_4[X; X_1, X_2]$ și afirmația este trivială.

Ca urmare a propoziției 7.6.4 rezultă că de fapt, odată cu coeficienții de corelație entropică (7.6.4) avem o infinitate de astfel de măsuri definite de relația (7.6.5), care coincid între ele valoric în cazul extrem al independenței sau în cel al reciproc complet dependenței.

7.6.1.2. Dependența a trei variabile aleatoare. Fie acum $X = (X_1, X_2, X_3)$ un vector aleator discret tridimensional și $H(X)$, $H(X_i)$, $H(X_i, X_j)$, $H(X_i/X_j)$, $H[H_i/(X_j, X_k)]$, $H[(X_i, X_j)/X_k]$, $1 \leqslant i \neq j \neq k \leqslant 3$, entropiile corespunzătoare acestuia definite în paragraful 3.3 Ca o generalizare naturală a relației (7.6.1) să considerăm expresia

$$W[X; X_1, X_2, X_3] = \sum_{j=1}^{3} H(X_j) - H(X). \qquad (7.6.6)$$

Analog definiției 7.6.1 vom spune că $W[X; X_1, X_2, X_3]$ se numește corelație entropică globală a variabilelor aleatoare X_1, X_2 și X_3. Aceasta exprimă gradul de interdependență sau de coeziune care există între X_1, X_2 și X_3 ca elemente constitutive ale lui X.

Ca urmare a relațiilor (3.3.19), (3.3.23), (3.3.24) și (3.3.25) putem scrie că pentru orice $1 \leqslant j \neq k \neq l \leqslant 3$

$$W[X; X_1, X_2, X_3] = H(X_j) + H(X_k) - H(X_k/X_l) - H[X_j/(X_k, X_l)] \qquad (7.6.7)$$

relație simetrică în raport cu j și $k \in \{1, 2, 3\}$ sau sub o altă formă

$$W[X; X_1, X_2, X_3] = H(X_j) + H(X_k) - H[(X_j, X_k)/X_l] \qquad (7.6.8)$$

Scanned with OKEN Scanner

Ținînd cont de proprietățile entropiilor corespunzătoare vectorului X și componentelor sale și în mod deosebit de propoziția 3.3.20, putem formula imediat

Propoziția 7.6.5. Pentru mărimea $W[X; X_1, X_2, X_3]$ sînt adevărate întotdeauna afirmațiile

1) este definită pentru orice vector aleator (X_1, X_2, X_3);
2) este simetică în raport cu variabilele X_1, X_2 și X_3;
3) $0 \leqslant W[X; X_1, X_2, X_3] < \infty$
4) $W[X; X_1, X, X_3] = 0$ dacă și numai dacă X_1, X_2 și X_3 sînt independente în totalitate;
5) $W[X; X_1, X_2, X_3] \leqslant H(X_j) + H(X_k)$, $1 \leqslant j \neq k \leqslant 3$, cu egalitate pentru o anumită pereche (j, k) dacă (X_j, X_k) c.d. X_i și cu egalitate pentru orice pereche (j, k) dacă și numai dacă vectorul X care componentele reciproc complet dependente în totalitate;
6) $W[X; X_1, X_2, X_3] \leqslant 2H(X_1, X_2, X_3]$ cu egalitate dacă și numai dacă X are componentele reciproc complet dependente în totalitate;
7) $W[\varphi(X); \varphi_1(X), \varphi_2(X), \varphi_3(X)] = W[X; X_1, X_2, X_3]$ pentru orice $\varphi = (\varphi_1, \varphi_2, \varphi_3)$, definite în mod biunivoc.

Ca o generalizare a propoziției 7.6.2 putem enunța

Propoziția 7.6.6. Pentru orice (X_1, X_2, X_3) avem

$$W[X; X_1, X_2, X_3] \leqslant \min\{H(X_j) + H(X_k)\} \leqslant$$
$$\leqslant \frac{2}{3} \sum_{j=1}^{3} H(X_j) \leqslant \max\{H(X_j) + HX_k)\} \leqslant 2H(X) \tag{7.6.9}$$

cu egalitate peste tot dacă și numai dacă X are componentele reciproc complet dependente în totalitate.

Demonstrația nu este dificilă, inegalitățile fiind, pe rînd, consecințe ale propoziției 3.3.20, înlănțuirea lor fiind dovedită imediat prin calcul direct.

Analog definiției 7.6.2, notînd cu $M_i[X_1, X_2, X_3]$ cei patru majoranți ai mărimii $W[X; X_1, X_2, X_3]$ din relația (7.6.9) în ordine descrescătoare, putem defini coeficienții de corelație entropică globală

$$G_i[X; X_1, X_2, X_3] = \frac{W[X; X_1, X_2, X_3]}{M_i[X_1, X_2, X_3]}, \quad 1 \leqslant i \leqslant 4 \tag{7.6.10}$$

pentru care putem formula cu ușurință două propoziții analoge propozițiilor 7.6.3 și 7.6.4

Scanned with OKEN Scanner

Să considerăm acum mărimea

$$W[X;(X_j, X_k), X_l] = H(X_j, X_k) + H(X_l) - H(X) \quad (7.6.11)$$

care apare ca o generalizare naturală a relației (7.6.1) și pe care o mai putem scrie și sub forma

$$W[X;(X_j, X_k), X_l] = H(X_j, X_k) - H[(X_j, X_k)|X_l] \quad (7.6.12)$$

sau sub forma

$$W[X;(X_j, X_k), X_l] = H(X_l) - H[X_l|(X_j, X_k)] \quad (7.6.13)$$

oricare ar fi $1 \leqslant j \neq k \neq l \leqslant 3$.

Definiția 7.6.3. Mărimea $W[X;(X_j, X_k), X_l]$ se numește corelație entropică a vectorului aleator (X_j, X_k) cu variabila aleatoare X_l, $1 \leqslant j \neq k \neq l \leqslant 3$.

Evident că $W[X;(X_j, X_k), X_l]$ exprimă tot un anumit grad al interdependenței dintre componentele lui X, grupate însă în (X_j, X_k) și X_l. Este vorba aici de a măsura interdependența dintre cele două părți constitutive ale lui X, respectiv vectorul (X_j, X_k) și variabila X_l.

Remarcăm că dacă X_j și X_k sînt independente, atunci relațiile (7.6.11) și (7.6.6) coincid și că în general, pentru orice $1 \leqslant j \neq k \neq l \leqslant 3$

$$W[X;X_1, X_2, X_3] \geqslant W[X;(X_j, X_k), X_l] \quad (7.6.14)$$

Ca urmare putem enunța

Propoziția 7.6.7. Are loc totdeauna egalitatea

$$W[X;X_1, X_2, X_3] =$$

$$= W[X;(X_l, X_k), X_l] + W[\hat{X};X_j, X_k] \quad (7.6.15)$$

unde $\hat{X} = (X_j, X_k)$, $1 \leqslant j \neq k \neq l \leqslant 3$.

Putem enunța acum o propoziție analogă propoziției 7.6.1 în care proprietățile 2), 5), 6) și 7) devin acum

2′) $W[X;(X_j;X_k), X_l] = W[X;X_l, (X_j, X_k)]$

5′) $W[X;(X_j, X_k), X_l] = H(X_j, X_k)$ dacă (X_j, X_k) c.d. X_l

5'') $W[X;\ (X_j,\ X_k),\ X_l] = H(X_l)$ dacă X_l c.d. $(X_j,\ X_k)$

6') $W[X;\ (X_j,\ X_k),\ X_l] = H(X_j, X_k) = H(X_l) = H(X)$ dacă și numai dacă (X_j, X_k) r.c.d. X_l

7') $W[\varphi(X);\ \varphi_1(X_j,\ X_k),\ \varphi_2(X_l)] = W[X;(X_j, X_k), X_l]$ pentru orice φ_1 și φ_2 definite în mod biunivoc.

Analog inegalităților (7.6.3) putem scrie că

$$W[X;\ (X_j,\ X_k),\ X_l] \leqslant \min\{H(X_j,\ X_k),\ H(X_l)\} \leqslant$$

$$\leqslant \frac{1}{2}[H(X_j,\ X_k) + H(X_l)] \leqslant$$

$$\leqslant \max\{H(X_j,\ X_k),\ H(X_l)\} \leqslant H(X) \qquad (7.6.16)$$

cu egalitate peste tot dacă și numai dacă (X_j, X_k) r.c.d. X_l pentru un anumit triplet (j, k, l) sau pentru orice $1 \leqslant j \neq k \neq l \leqslant 3$.

Ca urmare, putem defini coeficienții de corelație entropică dintre vectorul (X_j, X_k) și variabila X_l, $G_i[X;(X_j, X_k), X_l]$, $1 \leqslant i \leqslant 4$, analog relației (7.6.4) pentru care putem formula proprietățile similare din propoziția 7.6.3 cu modificările corespunzătoare propoziției 7.6.1, date mai sus. De asemenea putem formula propoziția 7.6.4. Ca o consecință imediată a relației (7.6.15), notînd pentru simplificare cu $G_i[X_1, X_2, X_3]$, $G_i[(X_j, X_k), X_l]$ și $G_i[X_j, X_k]$ cele trei categorii de coeficienți de corelație entropică, putem enunța

Propoziția 7.6.8. Pentru orice vector aleator (X_1, X_2, X_3) există numerele $0 \leqslant \alpha_i^{jk,l} \leqslant 1$ și $0 \leqslant \beta_i^{jk} \leqslant 1$ astfel încît

$$G_i[X_1, X_2, X_3] = \alpha_i^{jk,l}\, G_i[(X_j, X_k), X_l] + \beta_i^{jk}\, G_i[X_j, X_k]. \qquad (7.6.17)$$

Demonstrație. Ținînd cont de expresia fiecăruia din coeficienți deducem că

$$\alpha_i^{jk,l} = \frac{M_i[(X_j, X_k), X_l]}{M_i[X_1, X_2, X_3]} \text{ și } \beta_i^{jk} = \frac{M_i[X_j, X_k]}{M_i[X_1, X_2, X_3]} \qquad (7.6.18)$$

și afirmația este verificată.

Deși simplă, relația (7.6.17) este remarcabilă prin informația pe care o furnizează privind gradul de interdependență al tripletului $(X_j,\ X_k,\ X_l)$ obținut prin adăugarea lui X_l la cuplul (X_j, X_k), proprietate care apare ca naturală pentru

Scanned with OKEN Scanner

conceptul de dependenţă şi pe care o măsură a dependenţei trebuie să o reliefeze. Se poate constata cu uşurinţă că numai măsurile entropice ale dependenţei au această proprietate, de a putea realiza şi agregarea gradului de dependenţă, conform relaţiilor (7.6.15) şi (7.6.17).

7.6.1.3. Dependenţa a doi sau mai mulţi vectori aleatori. Să considerăm acum vectorul aleator discret n-dimensional $X = (X_1, X_2, \ldots, X_n)$ şi o partiţie a sa $Y = \{Y_j\}_{j=1}^{m}$, $m \leqslant n$, $Y_j \cap Y_k = \emptyset$, $j \neq k$, $X = \bigcup_{j=1}^{n} Y_j$. Să presupunem de asemenea cunoscute entropiile corespunzătoare $H(X)$, $H(X_j)$, $H(Y_k)$, $H(Y_k/Y_l)$, $1 \leqslant j \leqslant n$, $1 \leqslant k \neq 1 \leqslant m$, definite în paragraful 3.3.3. Ca urmare a propoziţiei 3.3.29 şi ca o generalizare naturală a relaţiei (7.6.11) să considerăm expresia

$$W[X; Y_1, \ldots, Y_m] = \sum_{j=1}^{m} H(Y_j) - H(X) \qquad (7.6.19)$$

pe care o vom numi corelaţie entropică globală a vectorilor $Y_1, \ldots, Y_m$ ca elemente ale partiţiei Y a lui X. Evident că

1) dacă $m = n = 2$, avem corelaţia a două variabile aleatoare, relaţia (7.6.19) coincizînd cu (7.6.1);
2) dacă $m = n = 3$, avem corelaţia a trei variabile aleatoare, relaţia (7.6.19) coincizînd cu (7.6.6);
3) dacă $n = 3$ şi $m = 2$, avem corelaţia dintre un vector bidimensinal şi o variabilă, relaţia (7.6.19) coincizînd cu (7.6.11);
4) dacă $m = n > 3$, avem corelaţia globală a n variabile aleatoare;
5) dacă $m < n > 3$, avem corelaţia globală a doi sau mai mulţi vectori aleatori.

Ca urmare a propoziţiei 3.3.29 deducem fără dificultate

Propoziţia 7.6.9. Următoarele afirmaţii asupra lui $W[X; Y_1, \ldots, Y_m]$ sînt adevărate întotdeauna.

1) este definită pentru orice X şi orice partiţie Y a lui X;
2) este simetrică în raport cu elementele partiţiei Y;
3) $0 \leqslant W[X; Y_1, \ldots, Y_m] < \infty$;
4) $W[X; Y_1, \ldots, Y_m] = 0$ dacă şi numai dacă vectorii $Y_1, \ldots, Y_m$ sînt independenţi în totalitate;
5) $W[X; Y_1, \ldots, Y_m] =$

$$= \sum_{k=2}^{m} \{H(Y_{j_k}) - H[Y_{j_k}/(Y_{j_1}, \ldots, Y_{j_{k-1}})]\} \quad 1 \leqslant j_k \leqslant m;$$

Scanned with OKEN Scanner

6) $W[X; Y_1 \ldots, Y_m] \leqslant \sum_{k=1}^{m-1} H(Y_{j_k})\ 1 \leqslant j_k \leqslant m$, cu egalitate dacă și numai dacă Y_{j_k} c.d. $[Y_{j_1}, \ldots, Y_{j_{k-1}}]$;

7) $W[X; Y_1, \ldots, Y_m] \leqslant (m-1)H(X)$ cu egalitate dacă și numai dacă $Y_1, \ldots, Y_m$ sînt reciproc complet dependenți;

8) $W[X; Y_1, \ldots, Y_m] \geqslant \sum_{k=2}^{m} [H(Y_{j_k}) - H(Y_{j_k}/Y_{j_{k-1}})]$;

9) este invariant față de orice transformare biunivocă a vectorilor $(Y_j)_{j=1}^m$;

10) $W[X; X_1, \ldots, X_n] \geqslant W[X; Y_1, \ldots, Y_m]$ pentru orice partiție Y a lui X, cu egalitate dacă $m = n$ sau în cazul independenței totale;

11) $W[X; X_1, \ldots, X_n \geqslant \sum_{j=1}^{m} W[Y_j]$, unde $W[Y_j]$ are forma (7.6.19), dar, pentru vectorul Y_j, $1 \leqslant j \leqslant m$, cu egalitate în cazul independenței globale.

Din proprietatea 5) rezultă

Propoziția 7.6.10. Dacă există o permutare $\{j_k\}_{k=1}^m$ a numerelor $\{1, 2, \ldots, m\}$ astfel încît Y_{j_k} c.d. cel puțin de unul din vectorii $Y_{j_1}, \ldots Y_{j_{k-1}}$, $2 \leqslant k \leqslant m$, atunci

$$W[X; Y_1, \ldots, Y_m] = \sum_{k=2}^{m} H(Y_{j_k}) \qquad (7.6.20)$$

și reciproc.

Propoziția 7.6.11. Dacă există o permutare $\{j_k\}_{k=1}^m$ a numerelor $\{1, 2, \ldots, m\}$ astfel încît $\{Y_{j_1}, Y_{j_2}, \ldots, Y_{j_{k-2}}\}$ c.d. $Y_{j_{k-1}}$, $3 \leqslant k \leqslant m$, atunci

$$W[X; Y_1, \ldots, Y_m] = \sum_{k=2}^{m} [H(Y_{j_k}) - H(Y_{j_k}/Y_{j_{k-1}})] \qquad (7.6.21)$$

și reciproc, iar dacă $\{Y_{j_1}, \ldots, Y_{j_{k-2}}\}$ c.d. $Y_{j_{k-1}}$ pentru orice permutare $\{j_k\}_{k=1}^m$, atunci

$$W[X; Y_1, \ldots, Y_m] = (m-1)H(Y_j) = \\ = (m-1)H(X), \quad 1 \leqslant j \leqslant m \qquad (7.6.22)$$

și reciproc.

Din proprietăţile 10) şi 11) rezultă că

1) dacă $X_1, \ldots, X_n$ sînt independente în totalitate, atunci orice vector Y_k şi ca urmare şi Y are componentele de asemenea independente în totalitate;
2) dacă cel puţin două componente ale lui X sînt dependente, atunci între componentele lui X există o anumită dependenţă globală.

Ţinînd cont de relaţia (3.3.34) deducem imediat

Propoziţia 7.6.12. Avem întotdeauna

$$W[X; Y_1, \ldots, Y_m] \leqslant \min \left\{ \sum_{k=1}^{m-1} H(Y_{j_k}) \right\} \leqslant$$

$$\leqslant \frac{m-1}{m} \sum_{j=1}^{m} H(Y_j) \leqslant \max \left\{ \sum_{k=1}^{m-1} H(Y_{j_k}) \right\} \leqslant (m-1)H(X) \tag{7.6.23}$$

cu egalitate peste tot dacă şi numai dacă $\{Y_j\}_{j=1}^m$ sînt reciproc complet dependente în totalitate.

Ca urmare a propoziţiei 3.3.27 rezultă că există numerele $0 \leqslant \alpha_{i_k} \leqslant 1$ astfel încît

$$W[X; X_1, \ldots, X_n] = \sum_{k=1}^{n} (1 - \alpha_{i_k}) H(X_{i_k}) \tag{7.6.24}$$

unde α_{i_k} reprezintă ponderea informaţiei proprii a lui X_{i_k} cînd se cunosc $X_{i_1}, X_{i_2}, \ldots, X_{i_{k-1}}$ în informaţia totală furnizată de X_{i_k}, $1 \leqslant k \leqslant n$. Deducem imediat

Propoziţia 7.6.13. Dacă $Z_1 = \{Y_j\}_{j \in J_1}$ şi $Z_2 = \{Y_j\}_{j \in J_2}$; $J_1 \cap J_2 = \emptyset$, $J_1 \cup J_2 = \{1, 2, \ldots, m\} = J$, atunci

$$W[X; Y_1, \ldots, Y_m] = W[Y; Z_1, Z_2] + W[Z_1; \{Y_j\}_{j \in J_1}] +$$

$$+ W[Z_2; \{Y_j\}_{j \in J_2}] \tag{7.6.25}$$

Ca urmare rezultă că pentru orice $Z = \{Y_j\}_{j \in J_1 \subset J}$, $W[Y] \geqslant W[Z]$ cu egalitate dacă şi numai dacă componentele lui Z sînt independente în totalitate şi $(Y - Z)$ este independent de Z.

Să notăm cu

$$W_k = W[\{Y_j\}_{j \in J_1}], \quad k = \operatorname{card} J_1, \quad 2 \leqslant k \leqslant m. \tag{7.6.26}$$

Scanned with OKEN Scanner

Rezultă imediat că $W_m \geqslant W_{m-1}$ cu egalitate dacă și numai dacă cel de al m-lea vector este independent de ceilalți $(m-1)$ vectori. Notînd acum cu $W_{1,k}$ corelația entropică dintre un vector și alți k vectori, atunci ca urmare a relației (7.6.25) rezultă că

$$W_m = W_{m-1} + W_{1,m-1} = \sum_{k=1}^{m-1} W_{1,k}. \qquad (7.6.27)$$

Relația (7.6.27) poate fi numită „relația de agregare" a dependenței cînd mulțimea vectorilor $Y_1, \ldots, Y_m$ s-a constituit prin juxtapunerea elementelor sale unul cîte unul.

Pentru generalizarea, de asemenea naturală, a relației (7.6.10) să considerăm expresiile $M_i[Y_1, \ldots, Y_m]$, $1 \leqslant i \leqslant 4$, majoranți în ordine descrescătoare ai corelației entropice din inegalitățile (7.6.23) și să notăm cu

$$G_i[X; Y_1, \ldots, Y_m] = \frac{W[X; Y_1, \ldots, Y_m]}{M_i[Y_1, \ldots, Y_m]} \quad 1 \leqslant i \leqslant 4 \qquad (7.6.28)$$

coeficienții de corelație entropică globală a vectorilor aleatori $Y_1, \ldots, Y_m$. Ca și în cazul corelației entropice globale, pentru diverse valori particulare ale lui m și n obținem coeficienți de corelație entropică globală pentru două sau mai multe variabile aleatoare, pentru doi sau mai mulți vectori aleatori sau corelația dintre variabile și vectori aleatori.

Ținînd cont de propozițiile 7.6.9—7.6.12 deducem

Propoziția 7.6.14. Următoarele afirmații asupra coeficienților de corelație entropică globală $G_i[X; Y_1, \ldots, Y_m]$ sînt adevărate întotdeauna

1) sînt definiți pentru orice vector aleator X și orice partiție y a acestuia, nici una din componente nefiind constantă cu probabilitatea unu;
2) sînt simetrici în raport cu vectorii $Y_1, Y_2, \ldots, Y_m$;
3) $0 \leqslant G_i[X; Y_1, \ldots, Y_m] \leqslant G_{i+1}[X; Y_1, \ldots, Y_m] \leqslant 1$, $1 \leqslant i \leqslant 3$;
4) $G_i[X; Y_1, \ldots, Y_m] = 0$, $1 \leqslant i \leqslant 4$, dacă și numai dacă $\{Y_j\}_{j=1}^m$ sînt independenți în totalitate;
5) $G_i[X; Y_1, \ldots, Y_m] = 1$, $1 \leqslant i \leqslant 4$, dacă și numai dacă $\{Y_j\}_{j=1}^m$ sînt reciproc complet dependenți în totalitate;
6) sînt invarianți față de orice transformare biunivocă a vectorilor $Y_1, \ldots, Y_m$.

Scanned with OKEN Scanner

Ca urmare a propoziției 7.6.9 mai putem deduce că

$$G_i[X; Y_1, \ldots, Y_m] \geqslant \frac{\sum_{k=2}^{m} [H(Y_{j_k}) - H(Y_{j_k}/Y_{k-1})]}{(m-1)\, H(X)}, \quad 1 \leqslant i \leqslant 4 \tag{7.6.29}$$

cu egalitate nulă dacă și numai dacă $\{Y_j\}_{j=1}^m$ sînt independenți în totalitate și cu egalitate nenulă dacă și numai dacă există o permutare $\{j_k\}_{k=1}^m$ a numerelor $\{1, 2, \ldots, m\}$ astfel încît $\{Y_{j_1}, \ldots, Y_{j_{k-2}}\}$ c.d. $Y_{j_{k-1}}$, $3 \leqslant k \leqslant m$.

Relația (7.6.29) constituie o completare a proprietății 3) din propoziția 7.6.14 și dă posibilitatea unei delimitări inferioare a gradului de dependență globală a vectorilor aleatori $\{Y_j\}_{j=1}^m$ în alte situații decît cele ale independenței, cînd egalitatea este zero, sau reciproc complet dependenței, cînd egalitatea este unu.

Propoziția 7.6.4 rămîne valabilă și pentru coeficienții de corelație entropică (7.6.28) și ca urmare, orice combinație liniară convexă a acestora este de asemenea un coeficient de corelație entropică.

Să notăm cu M_i^m și $M_i^{1,k}$ mărimile M_i din relația (7.6.28), calculate pentru vectorii $\{Y_j\}_{j=1}^m$ și pentru cuplul format dintr-un vector și alți k vectori. Ca urmare să notăm cu G_i^m și $G_i^{1,k}$ coeficienții de corelație entropică corespunzători. Ținînd cont și de relația (7.6.27) deducem

Propoziția 7.6.15. Pentru orice vectori aleatori $\{Y_j\}_{j=1}^m$ există numerele $\beta_i^{1,k} \in (0, 1)$, $\sum_{k=1}^{m-1} \beta_i^{1,k} > 1$, $1 \leqslant i \leqslant 4$, astfel încît

$$G_i^m = \sum_{k=1}^{m-1} \beta_i^{1,k} G_i^{1,k}. \tag{7.6.30}$$

Demonstrația este imediată, fiind suficient să luăm

$$\beta_i^{1,k} = \frac{M_i^{1,k}}{M_i^m} \quad 1 \leqslant k \leqslant m-1, \; 1 \leqslant i \leqslant 4. \tag{7.6.31}$$

Să remarcăm că relația (7.6.31) este o generalizare a relației (7.6.18). Ca și relația (7.6.27), dar sub altă formă, relația (7.6.31) este tot o „relație de agregare" a dependenței, atunci cînd familia de vectori aleatori $\{Y_j\}_{j=1}^m$ este obținută prin juxtapunerea acestora unul cîte unul, adică

$$\{Y_1, Y_2, \ldots, Y_m\} = (\ldots(((Y_1, Y_2), Y_3), Y_4), \ldots, Y_m). \tag{7.6.32}$$

Scanned with OKEN Scanner

Cantitativ, relația (7.6.31) demonstrează faptul că „gradul de dependență globală la nivelul ansamblului apare ca o sumă ponderată a gradelor de dependență ale acestuia corespunzătoare etapelor constituirii sale. Ținînd cont și de propoziția 7.6.4 în cazul general al dependenței, putem formula

Propoziția 7.6.16. Pentru orice $\alpha_i \geqslant 0$, $\sum_{i=1}^{4} \alpha_i = 1$, expresia

$$G[X; Y_1, \ldots, Y_m] = \sum_{i=1}^{4} \alpha_i \sum_{k=1}^{m-1} \beta_i^{1,k} G_i^{1,k} \qquad (7.6.33)$$

constituie de asemenea un coeficient de corelație entropică globală a vectorilor $\{Y_1, Y_2, \ldots, Y_m\}$.

Avînd în vedere propoziția 7.6.13, să notăm cu

$$W_{\text{tot}}[X] = \sum_{i=1}^{n} H(X_i) - H(X) \qquad \text{(a)}$$

$$W_{\text{tot}}[Y] = \sum_{j=1}^{m} H(Y_j) - H(Y) \qquad \text{(b)} \qquad (7.6.34)$$

$$W_{\text{tot}}[Y_j] = \sum_{i \in I_j} H(X_i) - H(Y_j) \qquad \text{(c)}$$

unde $Y_j = \{X_i\}_{i \in I_j}$, $I_j \cap I_k = \emptyset$, $\bigcup_{j=1}^{m} I_j = I = \{1, 2, \ldots, n\}$,

Rezultă imediat

Propoziția 7.6.17. Pentru orice vector aleator $X = (X_1, \ldots, X_n)$ și orice desfacere $Y = \{Y_j\}_{j=1}^{m}$ a acestuia

$$W_{\text{tot}}[X] = W_{\text{tot}}[Y] + \sum_{j=1}^{m} W_{\text{tot}}[Y_j]. \qquad (7.6.35)$$

Notînd cu $G_i[X]$, $G_i[Y]$ și $G_i[Y_j]$ coeficienții de corelație entropică globală corespunzători notației (7.6.34), din 7.6.35) rezultă că

$$G_i[X] = \sum_{j=0}^{m} \delta_i^j G_i[Y_j], \quad 1 \leqslant i \leqslant 4 \qquad (7.6.36)$$

unde $Y_0 = Y$ și $\delta_i^j = \dfrac{M_j[Y_j]}{M_i[X]}$, $0 \leqslant j \leqslant m$, $1 \leqslant i \leqslant 4$,

relația (7.6.36) constituind de asemenea „o relație de

Scanned with OKEN Scanner

agregare" a gradului de dependență sub forma cea mai generală.

7.6.2. Măsuri entropice ale dependenței unilaterale

Măsurile entropice ale dependenței globale furnizează informații prețioase asupra coeziunii sau interdependenței dintre două sau mai multe variabile aleatoare sau în cazul mai general dintre doi sau mai mulți vectori aleatori. Cu toate acestea ele nu pot spune în ce măsură depind unele de altele, deci care este gradul de dependență unilaterală dintre doi vectori aleatori sau în particular dintre două variabile aleatoare.

Fie X un vector aleator și

$$\mathfrak{D}(X) = \left\{ Y_j / Y_j \subset X,\ Y_j \cap Y_k = \emptyset,\ j \neq k,\ \bigcup_{k=1}^{m} Y_j \subseteq X \right\} \tag{7.6.37}$$

Definiția 7.6.4. O funcție $U : \mathfrak{D}(X) \times \mathfrak{D}(X) \setminus \{(Y_j, Y_j)\}_{j=1}^{m} \to [0,1]$ se numește măsură a dependenței unilaterale dintre vectorii aleatori Y_1 și Y_2 dacă

1) $U(Y_1, Y_2)$ există pentru orice cuplu (Y_1, Y_2) nici una din componente nefiind constantă cu probabilitatea unu
2) $U(Y_1, Y_2) \neq U(Y_2, Y_1)$ în general
3) $U(Y_1, Y_2) = U(Y_2, Y_1)$ dacă Y_1 r.c.d. Y_2 sau cînd Y_1 și Y_2 sînt independenți
4) $0 \leqslant U(Y_1, Y_2) \leqslant 1$
5) $U(Y_1, Y_2) = 0$ dacă și numai dacă Y_1 și Y_2 sînt independenți
6) $U(Y_1, Y_2) = 1$ dacă și numai dacă Y_1 c.d. Y_2
7) $U[\varphi_1(Y_1), \varphi_2(Y_2)] = U(Y_1, Y_2)$ pentru orice transformări reale biunivoc definite φ_1 și φ_2.

Să remarcăm că poprietățile sau axiomele impuse unei măsuri a dependenței unilaterale apar ca naturale ceea ce ne determină, ținînd cont și de postulatele lui Rényi, să afirmăm că în condițiile definiției 7.6.4, $U(Y_1, Y_2)$ poate fi considerată o bună măsură a dependenței unilaterale.

Propoziția 7.6.18. Mărimea

$$U(Y_1, Y_2) = \frac{H(Y_1) - H(Y_1/Y_2)}{H(Y_1)} \tag{7.6.38}$$

este o bună măsură a dependenței unilaterale a lui Y_1 față de Y_2.

Scanned with OKEN Scanner

Demonstrație. Axiomele 1) — 3) și 7) sînt evidente iar 4) — 6) rezultă din proprietățile entropiei condiționate.

Să observăm că dacă axioma 4 este formulată ca

$$4') \qquad 0 \leqslant U(Y_1, Y_2) < \infty$$

atunci axiomele 3) și 6) se exclud și ca urmare mărimea

$$U_1(Y_1, Y_2) = H(Y_1/Y_2) \tag{7.6.39}$$

constituie o măsură a dependenței unilaterale a lui Y_1 de Y_2. De asemenea dacă axioma 4) are forma

$$4'') \qquad 0 \leqslant U(Y_1, Y_2) < 1$$

atunci axiomele 3) și 6) sînt de asemenea excluse și mărimea

$$U_2(Y_1, Y_2) = \frac{H(Y_1/Y_2)}{H(Y_1, Y_2)} \tag{7.6.40}$$

constituie o măsură a dependenței globale. Remarcăm totodată că și complementara față de unu a mărimii (7.6.38) este o măsură a dependenței unilaterale satisfăcînd axiomele 1) — 7) dar cu 5) și 6) inversate ca posibilități de îndeplinire.

Definiția 7.6.5. În sensul definiției 7.6.4 spunem că Y_1 depinde de Y_2 mai mult decît Y_2 de Y_1 dacă $U(Y_1, Y_2) > U(Y_2, Y_1)$ și că Y_1 și Y_2 sînt egal unilateral dependenți dacă $U(Y_1, Y_2) = U(Y_2, Y_1)$.

Observăm imediat că dacă Y_1 și Y_2 sînt independenți sau Y_1 r.c.d. Y_2, atunci Y_1 și Y_2 sînt egal unilateral dependenți dar nu și reciproc. Ca urmare a relației (7.6.38) constatăm că

$$U(Y_1, Y_2) \leqslant U(Y_2, Y_1) \Leftrightarrow H(Y_1) \geqslant H(Y_2) \tag{7.6.41}$$

și ca urmare putem spune că Y_1 depinde mai mult de Y_2 decît Y_2 de Y_1 dacă $H(Y_1) < H(Y_2)$ și că Y_1 și Y_2 sînt egal unilateral dependenți dacă $H(Y_1) = H(Y_2)$.

Propoziția 7.6.19. Dacă $U(Y_1, Y_2)$ este dată de relația (7.6.38), atunci

1) $U(Y_1, Y_2) \geqslant G_1[Y_1, Y_2]$
2) $U(Y_1, Y_2) \leqslant U[Y_i(Y_2, Y_3)], \quad Y_j \in \mathfrak{D}(X), \quad i \leqslant j \leqslant 3$
3) există $0 < \alpha, \beta \leqslant \frac{1}{2}, \frac{1}{2} \leqslant \alpha + \beta \leqslant 1$ astfel că

$$G_1[Y_1, Y_2] = \alpha U(Y_1, Y_2) + \beta U(Y_2, Y_4) \tag{7.6.42}$$

Scanned with OKEN Scanner

unde $G_1[Y_1, Y_2]$ este coeficientul de corelație entropică globală $G_1[(Y_1, Y_2); Y_1, Y_2]$ dat de (7.6.28) în care $M_1[Y_1, Y_2] = H(Y_1, Y_2)$.

Demonstrație. Din relațiile (7.6.28) și (7.6.38) rezultă 1) și 3), unde $\alpha = \frac{H(Y_1)}{2H(Y_1, Y_2)}$ și $\beta = \frac{H(Y_2)}{2H(Y_1, Y_2)}$. Din definiția entropiei condiționate, rezultă că $H(Y_3/Y_2) \geqslant H[Y_3/(Y_1, Y_2)]$ cu egalitate dacă și numai dacă repartiția probabilistică a tripletului (Y_1, Y_2, Y_3) satisface condiția $p(y_1, y_2, y_3)p(y_2) = p(y_1, y_2)\, p(y_2, y_3)$ pentru orice (y_1, y_2, y_3), relație care este adevărată în cazurile extreme ale dependenței dintre Y_1, Y_2, Y_3 dar neimplicată de acestea. Ca urmare, deoarece

$$U[Y_1, (Y_2, Y_3)] = U(Y_1, Y_2) + \frac{H(Y_3/Y_2) - H[Y_3/(Y_1, Y_2)]}{H(Y_1)} \tag{7.6.43}$$

rezultă și punctul 2).

Să observăm că proprietatea 1) conduce la concluzia că gradul de dependență al unei componente a cuplului (Y_1, Y_2) de cealaltă componentă este mai mare decît gradul de interdependență a cuplului, iar proprietatea 2) arată că gradul de dependență al unui factor aleator de un grup de factori este mai mare decît gradul de dependență al acestuia de oricare dintre factorii grupului, proprietate care apare ca naturală și care ar putea fi luată chiar ca axiomă în definiția unei bune măsuri a dependenței unilaterale.

Să remarcăm acum că după cum Y_1 și Y_2 sînt vectori aleatori sau variabile aleatoare, putem vorbi de dependența unilaterală a unui vector de un alt vector sau de o variabilă precum și de cea a unei variabile de o altă variabilă sau de un vector. În oricare din cele patru cazuri posibile, măsura (7.6.38) rămîne neschimbată ca formă și la fel și celelalte două de altfel, numindu-le pe toate măsuri entropice ale dependenței unilaterale.

Din relația (7.6.43) deducem

$$U[Y_1, (Y_2, Y_3)] = U(Y_1, Y_j) + \alpha_{kj} U[Y_k/Y_j, Y_1] \tag{7.6.44}$$

unde $\alpha_{kj} = \frac{H(Y_k/Y_j)}{H(Y_1)}$, $2 \leqslant j \neq k \leqslant 3$ iar $U[Y_k/Y_j, Y_1]$, extinzînd într-un fel relația (7.6.38) exprimă gradul în care vectorul Y_k condiționat de vectorul Y_j depinde de vectorul Y_1, rezultat interesant în cazul în care Y_k și Y_j nu sînt

complet dependenţi unul de altul, cînd dealtfel $U[Y_1, (Y_2, Y_3)]$ şi $U(Y_1, Y_j)$ coincid, ceea ce spune că Y_1 depinde de cuplu ca de oricare din elementele sale.

7.6.3. Măsuri entropice ponderate ale dependenţei

Fie Y_1 şi Y_2 doi vectori aleatori şi să notăm cu

$$\Gamma[u; Y_1, Y_2] = -\sum_{y_1}\sum_{y_2} u(y_1, y_2)\, p(y_1, y_2) \log p(y_1, y_2) \quad (7.6.45)$$

entropia ponderată corespunzătoare cuplului (Y_1, Y_2), cu

$$\Gamma[u; Y_k] = -\sum_{y_1}\sum_{y_2} u(y_1, y_2)\, p(y_1, y_2) \log p(y_k) \quad (7.6.46)$$

entropia ponderată corespunzătoare vectorului Y_k, $1 \leqslant k \leqslant 2$ şi cu

$$\Gamma[u; Y_j/Y_k] = -\sum_{y_1}\sum_{y_2} u(y_1, y_2)\, p(y_1, y_2) \log p(y_j/y_k) \quad (7.6.47)$$

entropia ponderată corespunzătoare vectorului Y_j condiţionat de vectorul Y_k, $1 \leqslant j \neq k \leqslant 2$, pornind de la rezultatele din paragraful 4.2.

Din relaţia (7.6.46) deducem că

$$\Gamma[u; Y_1] = -\sum_{y_1} u(y_1)\, p(y_1) \log p(y_1) \quad (7.6.48)$$

unde

$$u(y_1) = \sum_{y_2} u(y_1, y_2)\, p(y_2/y_1) \quad (7.6.49)$$

şi analog putem scrie expresiile lui $\Gamma(u, Y_2]$ şi $u(y_2)$.

Ca urmare, deducem imediat

Propoziţia 7.6.20. Pentru orice cuplu aleator (Y_1, Y_2) avem

$$\Gamma[u; Y_1, Y_2] = \Gamma[u; Y_k] + \Gamma[u; Y_j/Y_k], \quad 1 \leqslant j \neq k \leqslant 2. \quad (7.6.50)$$

Ca o consecinţă imediată avem inegalitatea

$$\Gamma[u; Y_1, Y_2] \geqslant \Gamma[u; Y_k], \quad 1 \leqslant k \leqslant 2 \quad (7.6.51)$$

care devine egalitate dacă şi numai dacă $\Gamma[u; Y_j/Y_k] = 0$, $1 \leqslant j \neq j \leqslant 2$, ceea ce se poate întîmpla cînd $p(y_j/y_k) = 1$ pentru $(y_j, y_k) = (y_j^0, y_k^0)$ şi $p(y_j/y_k) = 0$ pentru $(y_j, y_k) \neq (y_j^0, y_k^0)$, deci cînd Y_1 şi Y_2 au o repartiţie comună cvasidiagonală, sau cînd $u(y_1, y_2) = 0$.

Scanned with OKEN Scanner

Vom spune că un experiment ce conduce la cuplul aleator (Y_1, Y_2) este de utilitate zero dacă $u(y_1, y_2) = 0$ pentru orice pereche (y_1, y_2) de valori posibile ale cuplului (Y_1, Y_2). În practică în general $u(y_1, y_2) \neq 0$ pentru cel puţin o pereche de valori ale cuplului aleator (Y_1, Y_2).

Din (7.6.46) şi (7.6.47) deducem că

$$\Gamma[u;\ Y_k] - \Gamma[u;\ Y_k/Y_j] =$$

$$= \sum_{y_1} \sum_{y_2} u(y_1, y_2)\, p(y_1, y_2) \log_2 \frac{p(y_1, y_2)}{p(y_1)\, p(y_2)} \qquad (7.6.52)$$

expresia (7.6.52) devenind zero dacă Y_1 şi Y_2 sînt independenţi presupunîndu-i de utilitate nenulă arbitrară. Pe de altă parte, ţinînd cont că $\log x \geqslant 1 - \frac{1}{x}$ cu egalitate dacă şi numai dacă $x = 1$, rezultă

$$\Gamma[u;\ Y_k] - \Gamma[u;\ Y_k/Y_j] \geqslant$$

$$\geqslant \sum_{y_1} \sum_{y_2} u(y_1, y_2)[p(y_1, y_2) - p(y_1)\, p(y_2)], \qquad (7.6.53)$$

de unde observăm de asemenea că dacă $u(y_1, y_2) =$ constant pentru orice cuplu (y_1, y_2) atunci $\Gamma[u;\ Y_k] \geqslant \Gamma[u;\ Y_k/Y_j]$ şi $\Gamma[u;\ Y_1, Y_2] \leqslant \Gamma[u;\ Y_1] + \Gamma[u;\ Y_2]$ pentru orice repartiţie a cuplului (Y_1, Y_2), cu egalitate în cazul independenţei.

Acelaşi rezultat se obţine din

Propoziţia 7.6.21. Avem întotdeauna

$$\Gamma[u;\ Y_k] - \Gamma[u;\ Y_k/Y_j] \geqslant$$

$$\geqslant \sum_{y_k} \sum_{y_j} \left[u(y_j, y_k)\, p(y_j, y_k) \cdot \log \frac{\sum\limits_{y_j} u(y_j, y_k)\, p(y_j, y_k)}{\sum\limits_{y_j} u(y_j, y_k)\, p(y_j)\, p(y_k)} \right] \qquad (7.6.54)$$

unde $1 \leqslant j \neq k \leqslant 2$.

Demonstraţie. Deoarece $f(x) = -x \log x$ este concavă, putem scrie inegalitatea lui Jensen conform căreia

$$f\left(\sum_{k=1}^{n} \lambda_k x_k \right) \geqslant \sum_{k=1}^{n} \lambda_k f(x_k), \quad \lambda_k \geqslant 0, \ \sum_{k=1}^{n} \lambda_k = 1.$$

Scanned with OKEN Scanner

Prin urmare să luăm

$$\lambda_k = \frac{u(y_j, y_k)\, p(y_k)}{\sum\limits_{y_k} u(y_j, y_k) p(y_k)} \quad \text{și} \quad x_k = p(y_j/y_k). \qquad (7.6.55)$$

Rezultă

$$-\sum_{y_k} \frac{u(y_j, y_k)\, p(y_j, y_k)}{\sum\limits_{y_k} u(y_j, y_k)\, p(y_k)} \log \frac{\sum\limits_{y_k} u(y_j, y_k)\, p(y_j, y_k)}{\sum\limits_{y_k} u(y_j, y_k)\, p(y_k)} \geqslant$$

$$\geqslant -\sum_{y_k} \frac{u(y_j, y_k)\, p(y_j, y_k)}{\sum\limits_{y_k} u(y_j, y_k)\, p(y_k)} \log_p(y_j/y_k) \qquad (7.6.56)$$

Simplificînd cu $\sum\limits_{y_k} u(y_j, y_k)\, p(y_k)$ și însumînd apoi și după y_j rezultă

$$-\sum_{y_j}\sum_{y_k} u(y_j, y_k)\, p(y_j, y_k) \log \frac{\sum\limits_{y_k} u(y_j, y_k)\, p(y_j, y_k)}{\sum\limits_{y_k} u(y_j, y_k)\, p(y_k)} \geqslant$$

$$\geqslant -\sum_{y_j}\sum_{y_k} u(y_j, y_k)\, p(y_j, y_k) \log_p(y_j/y_k) \qquad (7.6.57)$$

ceea ce, ținînd cont de (7.6.47) reprezintă un majorant pentru $\Gamma[u\,;\, Y_j/Y_k]$. Permutînd j cu k obținem un majorant pentru $\Gamma[u\,;\, Y_k/Y_j]$ și apoi scăzînd $\Gamma[u\,;\, Y_k/Y_j]$ din $\Gamma[u, Y_k]$ obținem (7.6.54), după un calcul simplu în care ținem cont și de (7.6.46).

Ținînd cont de cele de mai sus putem considera cu ușurință expresia

$$W[u\,;\, Y_1, Y_2] = \Gamma[u\,;\, Y_1] + \Gamma[u\,;\, Y_2] - \Gamma[u\,; Y_1, Y_2] \qquad (7.6.58)$$

care de altfel reprezintă, ca și în cazul neponderat, balanța informațională

$$W[u\,;\, Y_1, Y_2] = \Gamma[u\,;\, Y_j] - \Gamma[u\,;\, Y_j/Y_k]$$

$$1 \leqslant j \neq k \leqslant 2. \qquad (7.6.59)$$

Analog rezultatelor din paragraful 7.6.1 putem da acum

Definiția 7.6.6. Se numește corelație entropică ponderată a vectorilor aleatori Y_1 și Y_2 mărimea $W[u\,;\, Y_1, Y_2]$ dată de relația (7.6.58).

Scanned with OKEN Scanner

Să remarcăm că $W[u; Y_1, Y_2]$ exprimă intensitatea coeziunii ponderate sau gradul de interdependență ponderată dintre Y_1 și Y_2 în sensul că se au în vedere nu numai probabilitățile de realizare a evenimentelor ce definesc cei doi vectori aleatori ci și importanța sau utilitatea realizării acestor evenimente. Prezența utilității în expresia unei măsuri a interdependenței poate conduce la valori mai mari sau mai mici ale mărimii $W[u; Y_1, Y_2]$ față de situația în care utilitatea nu se ia în seamă, ceea ce în formula (7.6.58) este echivalent cu $u(y_1, y_2) = 1$ pentru orice pereche de valori (y_1, y_2) ale cuplului aleator (Y_1, Y_2). Ca urmare a acestui fapt se poate spune că prezența utilității în expresia unei măsuri a dependenței probabiliste reprezintă o pondere a interdependenței probabiliste dintre Y_1 și Y_2. Efectul ponderării este ușor vizibil în însăși faptul că măsurile entropice ponderate nu mai au aceleași proprietăți ca și măsurile entropice neponderate. Un prim exemplu îl constituie chiar entropia ponderată.

Propoziția 7.6.22. Corelația entropică ponderată satisface următoarele:

1) este definită pentru orice cuplu aleator (Y_1, Y_2) și orice ponderi $\{u(y_1, y_2)\}_{y_1, y_2}$ ale acestuia;

2) $W[u; Y_1, Y_2] = W[u; Y_2, Y_1]$ pentru orice ponderi $\{u(y_1, y_2)\}_{y_1, y_2}$ ale lui (Y_1, Y_2);

3) $\alpha_s \leqslant W[u; Y_1, Y_2] < \infty$, $s = 1, 2$, unde α_1 este minorantul din (7.6.53), iar α_2 minorantul din (7.6.54) pentru $W[u; Y_1, Y_2]$;

4) dacă orice două rezultate (y_1, y_2) sînt egal ponderate, atunci $W[u; Y_1, Y_2] \geqslant 0$;

5) dacă Y_1 și Y_2 sînt independenți probabilistic atunci $W[u; Y_1, Y_2] = 0$, oricare ar fi ponderile $\{u(y_1, y_2)\}_{y_1, y_2}$;

6) dacă pentru fiecare valoare a lui y_k există un y_j^k astfel încît $p(y_j^k/y_k) = 1$ și $p(y_j/y_k) = 0$ dacă $y_j \neq y_j^k$, $1 \leqslant j \neq k \leqslant 2$ atunci

$$W[u; Y_1, Y_2] = \Gamma[u; Y_1] = \Gamma[u; Y_2] = \\ = \Gamma[u; Y_1, Y_2]; \qquad (7.6.60)$$

7) pentru un sistem de ponderi $\{u(y_1, y_2)\}_{y_1, y_2}$ dat, $W[u; Y_1, Y_2]$ este invariantă față de orice transformare biunivocă a vectorilor Y_1 și Y_2.

Din relațiile (7.6.51) și (7.6.52) deducem cu ușurință

Scanned with OKEN Scanner

Propoziția 7.6.23. Avem întotdeauna

$$W[u;\ Y_1,\ Y_2] \leqslant \min_{j=1,\,2}\{\Gamma[u;\ Y_j]\} \leqslant \frac{1}{2}[\Gamma[u;\ Y_1]+$$

$$+\ \Gamma[u;\ Y_2]] \leqslant \max_{j=1,\,2}\{\Gamma[u;\ Y_j]\} \leqslant \Gamma[u;\ Y_1,\ Y_2] \qquad (7.6.61)$$

cu egalitate peste tot în condițiile relației (7.6.60).

Să remarcăm că dacă $u(y_1, y_2) = 1$ pentru orice pereche de valori (y_1, y_2) ale cuplului (Y_1, Y_2) atunci relațiile (7.6.61) și (7.6.3) coincid cînd Y_1 și Y_2 sînt variabile aleatoare.

Analog rezultatelor din pragraful 7.6.1 putem introduce acum coeficienții de corelație entropică ponderată

$$G_i[u;\ Y_1,\ Y_2] = \frac{W[u;\ Y_1,\ Y_2]}{M_i[u;\ Y_1,\ Y_2]} \qquad 1 \leqslant i \leqslant 4 \qquad (7.6.62)$$

unde $M_i[u;\ Y_1,\ Y_2]$, $1 \leqslant i \leqslant 4$, sînt majoranții în ordine descrescătoare din relația (7.6.61), ceea ce conduce la relația $G_i[u;\ Y_1,\ Y_2] \leqslant G_{i+1}[u;\ Y_1,\ Y_2]$ pentru $1 \leqslant i \leqslant 3$. Să presupunem acum că ponderile $u(y_1, y_2)$ nu sînt toate nule.

Definiția 7.6.7. Vom spune că Y_1 este ponderat complet dependent de Y_2 dacă $\Gamma[u;\ Y_1/Y_2] = 0$ și că Y_1 și Y_2 sînt ponderat reciproc complet dependenți dacă $\Gamma[u;\ Y_1/Y_2] = \Gamma[u;\ Y_2/Y_1] = 0$.

Definiția 7.6.8. Vom spune că Y_1 și Y_2 sînt entropic ponderat independenți dacă $\Gamma[u;\ Y_1,\ Y_2] = \Gamma[u;\ Y_1] + \Gamma[u;\ Y_2]$.

Remarcăm că dacă Y_1 și Y_2 sînt probabilistic independenți atunci ei sînt și entropic ponderați independenți. Ținînd cont de cele două definiții, proprietățile 5) și 6) din propoziția 7.6.22 pot fi formulate în acest sens. Ca urmare putem formula

Propoziția 7.6.24. Pentru coeficienții de corelație entropică ponderată (7.6.62) sînt adevărate următoarele

1) sînt definiți pentru orice cuplu aleator neconstant și de utilitate nenulă;

2) $G_i[u;\ Y_1,\ Y_2] = G_i[u;\ Y_2,\ Y_1]$, $1 \leqslant i \leqslant 4$, pentru orice sistem de utilități $\{u(y_1, y_2)\}_{y_1, y_2}$;

3) $\beta_s^i \leqslant G_i[u;\ Y_1,\ Y_2] \leqslant 1$; $1 \leqslant i \leqslant 4$, $s = 1, 2$, unde $\beta_s^i = \dfrac{\alpha_s}{M_i[u;\ Y_1,\ Y_2]}$ și α_s din propoziția 7.6.22;

4) dacă $u(y_1, y_2) =$ constant oricare ar fi (y_1, y_2) atunci

$$G_i[u;\ Y_1,\ Y_2] \geqslant 0, \quad 1 \leqslant i \leqslant 4;$$

Scanned with OKEN Scanner

5) $G_i[u; Y_1, Y_2] = 0$, $1 \leqslant i \leqslant 4$, dacă și numai dacă Y_1 și Y_2 sînt entropic ponderat independenți;

6) $G_i[u; Y_1, Y_2] = 1$, $1 \leqslant i \leqslant 4$, dacă și numai dacă Y_1 și Y_2 sînt ponderat reciproc complet dependenți;

7) pentru un anumit sistem de utilități sau ponderi $\{u(y_1, y_2)\}_{y_1, y_2}$ $G_i[u; Y_1, Y_2]$, $1 \leqslant i \leqslant 4$, sînt invarianți în raport cu orice transformare biunivocă a vectorilor aleatori Y_1 și Y_2.

Deducem, analog propoziției 7.6.4,

Propoziția 7.6.25. Orice combinație liniară convexă

$$G[u; Y_1, Y_2; \alpha] = \sum_{i=1}^{4} \alpha_i G_i[u; Y_1, Y_2],$$

$$\alpha_i \geqslant 0, \sum_{i=1}^{4} \alpha_i = 1 \qquad (7.6.63)$$

a coeficienților de corelație entropică ponderată este de asemenea un coeficient de corelație entropică ponderată.

Să observăm acum, cu titlu de exemplu, că dacă Y_1 și Y_2 sînt două variabile aleatoare discrete pozitive, atunci

a) dacă $u(y_1, y_2) = y_1 \cdot y_2$ rezultă că

$$W[u; Y_1, Y_2] = \sum_{y_1} \sum_{y_2} y_1 y_2 p(y_1, y_2) \log \frac{p(y_1, y_2)}{p(y_1)\, p(y_2)}; \qquad (7.6.64)$$

b) dacă $u(y_1, y_2) = \dfrac{y_1 y_2 \operatorname{sign}\left[\log \dfrac{p(y_1, y_2)}{p(y_1)\, p(y_2)}\right]}{\log \dfrac{p(y_1, y_2)}{p(y_1)\, p(y_2)}}$, rezultă că

$$W[u; Y_1, Y_2] = M[Y_1 \cdot Y_2] \qquad (7.6.65)$$

adică tocmai valoarea medie a variabilei aleatoare produs $Y_1 \cdot Y_2$.

Revenind la relația (7.6.58), în același mod ca în cazul neponderat putem introduce acum entropia ponderată

$$\Gamma[u; Y] = \Gamma[u; Y_1, \ldots, Y_m] =$$

$$= -\sum_{y_1} \cdots \sum_{y_m} u(y_1, \ldots, y_n)\, p(y_1, \ldots, y_m) \cdot \log p(y_1, \ldots, y_m) \qquad (7.6.66)$$

și ca urmare, definim corelația entropică ponderată globală a vectorilor aleatori $Y_1, \ldots, Y_m$ ca fiind

$$W[u; Y_1, \ldots, Y_m] = \sum_{j=1}^{m} \Gamma[u; Y_j] - \Gamma[u; Y] \quad (7.6.67)$$

și coeficienții de corelație entropică ponderată globală $G_i[u; Y_1, \ldots, Y_m]$, $1 \leqslant i \leqslant 4$, generalizînd astfel și relația (7.6.62).

Pentru mărimile (7.6.66) și (7.6.67) putem prezenta unele rezultate analoge celor din paragraful 7.6.1.3, corespunzătoare măsurilor dependenței globale a vectorilor aleatori $Y_1, Y_2, \ldots, Y_m$ în cazul neponderat. Nu rămîn toate valabile întrucît mărimile (7.6.19) și (7.6.28) sînt cazuri particulare ale mărimilor (7.6.66) și (7.6.67).

Prin analogie cu expresia (7.6.38) putem introduce și o măsură a dependenței unilaterale ponderate considerînd expresia

$$U[u; Y_1, Y_2] = \frac{|\Gamma[u; Y_1] - \Gamma[u; Y_1/Y_2]|}{\Gamma[u; Y_1]} \quad (7.6.68)$$

pentru care putem formula proprietăți asemănătoare celor din definiția 7.6.4. Luarea în valoare absolută a raportului (7.6.68) se face pentru a avea, indiferent de ponderi, $0 \leqslant$ $\leqslant U[u; Y_1, Y_2] \leqslant 1$.

7.6.4. Conexiuni ale măsurilor entropice cu alte măsuri ale dependenței

Fie $X = (X_1, X_2)$ un vector aleator continuu a cărui repartiție este dată de relația (7.2.1). Ca urmare a relațiilor (7.2.2) — (7.2.4) rezultă

Propoziția 7.6.26. Dacă $X = (X_1, X_2)$ urmează repartiția (7.2.1) atunci

$$W[X; X_1, X_2] = -\frac{1}{2}\log_2[1 - \rho^2(X_1, X_2)]. \quad (7.6.69)$$

Se observă imediat că dacă

$$a_k = \frac{1}{\sigma_k^2(1 - \rho^2)}, \quad k = 1, 2; \quad h = -\frac{\rho}{\sigma_1\sigma_2(1 - \rho^2)} \quad (7.6.70)$$

atunci repartiția (7.2.1) este chiar repartiția normală bidimensională de medii nule, dispersii σ_1^2 și σ_2^2 și coeficient de

Scanned with OKEN Scanner

corelație ρ^2. Relația (7.6.69) rămîne neschimbată în cazul particular (7.6.70), ceea ce dovedește că între $W[X; X_1, X_2]$ și $\rho(X_1, X_2)$ există o legătură funcțională mai puțin restrictivă decît cea cerută de postulatul al șaptelea al lui Rényi, postulat care poate fi satisfăcut însă în versiunea lui C. B. Bell (1962), căci $W[X; X_1, X_2]$ ca funcție de ρ, este strict monotonă.

Să remarcăm de asemenea fără dificultate că pentru repartiția (7.6.69) și cazul particular (7.6.70) al acesteia avem :

$$H(X_1, X_2) = \log 2\pi e [\sqrt{a_1 a_2 - h^2}]^{-1} =$$

$$= \log 2\pi e \sigma_1 \sigma_2 \sqrt{1 - \rho^2} \qquad (7.6.71)$$

$$H(X_k) = \log \sqrt{2\pi e a_j} [\sqrt{a_1 a_2 - h^2}]^{-1} =$$

$$= \log \sigma_k \sqrt{2\pi e}; \quad 1 \leqslant k \neq j \leqslant 2. \qquad (7.6.72)$$

Ca urmare, mărimile $M_i[X_1, X_2]$ din relația (7.6.4) devin, în cazul (7.6.70),

$$M_1[X_1, X_2] = \log 2\pi e \sigma_1 \sigma_2 \sqrt{1 - \rho^2};$$

$$M_2[X_1, X_2] = \max_k \{\log \sigma_k \sqrt{2\pi e}\};$$

$$M_3[X, X_1] = \log 2\pi e \sqrt{\sigma_1 \sigma_2};$$

$$M_4[X_1, X_2] = \min_k \{\log \sigma_k \sqrt{2\pi e}\}$$

ceea ce permite o explicitare a coeficienților de corelație entropică $G_i[X; X_1, X_2]$ în funcție de ρ, $1 \leqslant i \leqslant 4$. Se poate constata cu ușurință că $G_i[X; X_1, X_2]$ sînt funcții de ρ strict monotone, în anumite cazuri particulare chiar cu același tip de monotonie. În consecință putem formula.

Propoziția 7.6.27. Coeficienții de corelație entropică $G_i[X; X_1, X_2]$ definiți de relația (7.6.4) constituie „bune măsuri" ale dependenței în sensul lui Rényi-Bell.

Ca urmare a propozițiilor 7.6.4 și 7.6.27 putem spune chiar că orice combinație liniară convexă a coeficienților de corelație entropici constituie o „bună măsură" a dependenței

Scanned with OKEN Scanner

în sensul lui Rényi-Bell. Din relațiile (7.6.4) și (7.6.69) deducem că pentru repartiția (7.2.1) avem

$$\rho^2(X_1, X_2) = 1 - e^{-\alpha_i G_i[X; X_1, X_2]} \tag{7.6.73}$$

unde $\alpha_i = 2M_i[X_1, X_2]$, $1 \leqslant i \leqslant 4$, ceea ce sugerează considerarea expresiei

$$\delta(X_1, X_2) = \sqrt{1 - e^{-\alpha_i G_i[X; X_1, X_2]}} \tag{7.6.74}$$

care nu este altceva decît coeficientul de corelație informațională al lui Linfoot (1957) și care pentru $i = 2$ și $i = 4$ apare ca o funcție de coeficienții de corelație informațională ai lui Bell (1962).

Deoarece $\alpha_i G_i[X; X_1, X_2] = 2W[X; X_1, X_2]$, relația (7.6.74) definește o singură măsură a dependenței dintre X_1 și X_2, care în cazul repartiției (7.6.70) coincide cu $|\rho(X_1, X_2)|$ și care satisface toate postulatele lui Rényi.

Relația (7.6.74) sugerează posibilitatea unei extensiuni a postulatelor lui Rényi pentru mai mult de două variabile.

Fie $X = (X_1, \ldots, X_n)$ un vector aleator continuu a cărui repartiție este dată de relația (3.4.9) și pentru care se obțin expresiile

$$H(X) = \frac{1}{2} \log \frac{(2\pi e)^n}{|A|} \text{ și } H(X_i) =$$

$$= \frac{1}{2} \log \frac{2\pi e\, M_{ii}}{|A|}, \quad 1 \leqslant i \leqslant n \tag{7.6.75}$$

și ca urmare

$$W[X; X_1, \ldots, X_n] = \frac{1}{2} \log \frac{\prod_{i=1}^{n} M_{ii}}{|A|^{n-1}} \tag{7.6.76}$$

Se poate arăta că dacă matricea $A = (a_{ij})$ care apare în repartiția (3.4.9) este simetrică, reală și pozitiv definită, astfel încît $a_{ii} a_{jj} - a_{ij}^2 \geqslant 0$ și $M_{ii} > 0$ pentru orice $1 \leqslant i, j \leqslant n$, atunci

$$|A|^{n-1} \leqslant \prod_{i=1}^{n} M_{ii} \tag{7.6.77}$$

cu egalitate dacă și numai dacă $a_{ij} = 0$ pentru orice $1 \leqslant i \neq j \leqslant n$, caz care înseamnă că X_i și X_j sînt independente,

Scanned with OKEN Scanner

repartiţia (3.4.9) apărînd atunci ca produs al repartiţiilor individuale. Ca urmare să considerăm mărimea

$$r^2(X_1, \ldots, X_n) = 1 - \frac{|A|^{n-1}}{\prod_{i=1}^{n} M_{ii}} \qquad (7.6.78)$$

pentru care putem observa că :

1) este definită pentru orice vector aleator n-dimensional X ;
2) este simetrică în raport cu $X_1, X_2, \ldots, X_n$;
3) $0 \leqslant r(X_1, \ldots, X_n) \leqslant 1$;
4) $r(X_1, \ldots, X_n) = 0$ dacă şi numai dacă X are componentele independente două cîte două ;
5) $r(X_1, \ldots, X_n) = 1$ dacă şi numai dacă $|A| = 0$, adică atunci şi numai atunci cînd X are componentele dependente unele de altele ;
6) este invariantă faţă de orice transformare biunivocă a componentelor vectorului X ;
7) dacă $n = 2$ atunci $r^2(X_1, X_2) = \rho^2(X_1, X_2)$.

Acceptînd pe $r(X_1, X_2, \ldots, X_n)$ ca măsură a dependenţei globale a variabilelor $X_1, X_2, \ldots, X_n$, atunci, ţinînd cont de relaţiile (7.6.76) şi (7.6.77) am putea considera mărimea

$$\delta^2(X_1, \ldots, X_n) = 1 - e^{-2\,W[X; X_1, \ldots, X_n]} \qquad (7.6.78)$$

ca o extensie naturală a relaţiei (7.6.74) şi cu aceasta am putea reformula postulatele lui Rényi pentru cazul dependenţei a n variabile aleatoare şi apoi mai general pentru m vectori aleatori, cu precizarea formei lui Bell pentru cel de-al şaptelea postulat al lui Rényi.

Ţinînd cont de relaţiile (7.3.5) şi (7.6.1) putem deduce imediat

Propoziţia 7.6.28. Dacă $\Omega = \mathbf{R}^2$, iar ipotezele H_1 şi H_2 asupra lui X conduc la relaţiile (7.3.18) şi (7.3.19) atunci

$$W[X ; X_1, X_2] = I[1 : 2 ; X_1, X_2]. \qquad (7.6.80)$$

Ţinînd cont de proprietăţile J-divergenţei deducem de asemenea imediat

Propoziţia 7.6.29. În condiţiile propoziţiei 7.6.28 între corelaţia entropică, J-divergenţă şi media pătratului de contingenţă au loc relaţiile

$$W[X ; X_1, X_2] \leqslant J(1, 2) \leqslant C^2(X_1, X_2) \qquad (7.6.81)$$

Scanned with OKEN Scanner

cu prima egalitate dacă şi numai dacă X_1 şi X_2 sînt independente şi cu a doua dacă X_1 şi X_2 sînt independente sau sînt normal repartizate, caz în care $J(1, 2) = C^2(X_1, X_2) = \frac{\rho^2}{1 - \rho^2}$.

Inegalitatea (7.6.81) apare ca un argument în plus pentru a considera J-divergenţa lui Kullback drept o măsură a dependenţei.

Fie acum α şi $\beta \in [0, \infty)$. Se poate verifica uşor că dacă $0 \leqslant \alpha \leqslant \beta$ atunci $\frac{\alpha}{\alpha+1} \leqslant \frac{\beta}{\beta+1}$ cu egalitate dacă şi numai dacă $\alpha = \beta$, iar dacă $0 \leqslant \alpha \leqslant \frac{1}{2} \ln(1+\beta)$, atunci $1 - e^{-2\alpha} \leqslant \frac{\beta}{1+\beta}$ cu egalitate dacă şi numai dacă $\alpha = \frac{1}{2}\ln(1+\beta)$. Ca urmare a acestui fapt deducem imediat

Propoziţia 7.6.30. În condiţiile propoziţiei 7.6.28 avem

$$1 - e^{-2W[X; X_1, X_2]} \leqslant \bar{J}(1, 2) \leqslant C_1^2(X_1, X_2) \qquad (7.6.82)$$

cu egalitate peste tot dacă X_1 şi X_2 sînt independente sau normal repartizate, caz în care cele trei mărimi devin $\rho^2(X_1, X_2)$.

Remarcăm că relaţia (7.6.82) este de fapt relaţia (7.6.81) formulată însă pentru cele trei măsuri ale dependeţei modificate astfel încît să satisfacă postulatul al şaptelea al lui Rényi.

Scanned with OKEN Scanner

III. INFORMAȚIE ȘI CORELAȚIE ENTROPICĂ ÎN TEORIA SISTEMELOR

8. Interdependența

8.1. Sisteme

Ca în multe alte domenii științifice sau de altă natură și în teoria sistemelor există o multitudine de sensuri pentru unul și același lucru, ceea ce impune în mod necesar unele precizări asupra conceptelor utilizate.

Vom numi sistem, în sensul cel mai larg, un tot unitar, rezultat din unirea mai multor elemente de același gen, între care există o anumită interdependență și care constituie astfel un tot organizat. Orice parte a unui sistem se numește subsistem. Spunem că o anumită parte a unui sistem este indivizibilă dacă aceasta nu poate fi socotită, sau nu dorim să fie socotită la rîndul ei ca un sistem. Un element constitutiv indivizibil al unui sistem se numește atom sau bloc. Un sistem care nu conține nici un bloc se numește sistem vid. Dacă existența sistemului este obiectivă, se spune că sistemul este real. Dacă indiferent de momentul la care se fac observații asupra lui, sistemul este același, se spune că avem un sistem staționar, iar în caz contrar că avem un sistem evolutiv. Dacă la fiecare moment se poate cunoaște cu precizie starea în care se află sistemul, se spune că evoluția sa este deterministă sau că sistemul este determinist, iar dacă sistemul se află într-o anume stare cu o anumită probabilitate, atunci se spune că avem un sistem aleator. Raporturile care există între elementele unui sistem în timpul evoluției acestuia se numesc conexiuni sau interdependențe. După cum mulțimea blocurilor constitutive este finită, numărabilă sau infinită, vom spune că sistemul este finit, numărabil sau infinit.

Vom nota un sistem cu $\mathcal{S}$, un subsistem al său cu $\mathcal{T}$ și un bloc cu B, scriind ca de obicei în teoria mulțimilor, $\mathcal{T} \subseteq \mathcal{S}$, $\mathcal{B} \in \mathcal{S}$, $\mathcal{S} = \{B_i\}_{i \in I}$, $\mathcal{T} = \{B_i\}_{i \in I_1 \subseteq I}$, unde I este mulțimea indicilor blocurilor lui $\mathcal{S}$. Vom nota cu $\mathfrak{P}(\mathcal{S}) = \{\mathcal{T} / \mathcal{T} \subseteq \mathcal{S}\}$

Scanned with OKEN Scanner

mulțimea tuturor subsistemelor lui $\mathcal{S}$, inclusiv subsistemul vid $\emptyset$ și sistemul însuși $\mathcal{S}$, iar cu $Q[\mathcal{S}]$ mulțimea tuturor conexiunilor dintre blocurile lui $\mathcal{S}$. Evident că dacă $\mathcal{S}$ este finit și conține n blocuri, atunci $Q[\mathcal{S}]$ conține $2^n - -(n+1)$ conexiuni. Vom nota cu $\mathfrak{D}[\mathcal{S}]$ o desfacere a lui $\mathcal{S}$ și anume

$$\mathfrak{D}[\mathcal{S}] = \{\mathcal{T}_j\}_{j\in J},\ \mathcal{T}_j \cap \mathcal{T}_k = \emptyset,\quad j \neq k \quad \bigcup_{j\in J} \mathcal{T}_j = \mathcal{S} \qquad (8.1.1)$$

Evident că dacă $J = I$, atunci $\mathfrak{D}[\mathcal{S}] = \mathcal{S}$, caz în care vom spune că avem o desfacere atomică a lui $\mathcal{S}$. Corespunzător desfacerii $\mathfrak{D}$ a lui $\mathcal{S}$ vom nota cu $Q_{\mathfrak{D}}[\mathcal{S}]$ mulțimea conexiunilor dintre subsistemele $\mathcal{T}_j$, $j \in J$, considerate ca părți indivizibile ale lui $\mathcal{S}$, conexiuni numite și conexiuni specifice desfacerii $\mathfrak{D}[\mathcal{S}]$. Dacă $\mathcal{S}$ este format din n blocuri, iar $\mathfrak{D}[\mathcal{S}]$ conține m subsisteme ale lui $\mathcal{S}$, fiecare cu cîte n_k blocuri, $\sum_{k=1}^{m} n_k = n$, atunci, la nivelul familiei de subsisteme $\mathfrak{D}[\mathcal{S}]$ putem vorbi de conexiuni între subsistemele lui $\mathcal{S}$, între blocurile fiecărui subsistem, precum și de conexiuni între blocuri ale unui subsistem cu blocuri ale altor subsisteme sau chiar cu subsisteme. Desigur, nu toate aceste conexiuni prezintă interes. Dar dacă un anumit studiu urmărește cunoașterea tuturor conexiunilor posibile, numărul lor este suficient de mare, iar în contextul considerării unei desfaceri $\mathfrak{D}[\mathcal{S}]$ acest număr sporește și mai mult. Dacă $\mathfrak{D}[\mathcal{S}] = \mathcal{S}$, atunci $Q_{\mathfrak{D}}[\mathcal{S}] = Q[\mathcal{S}]$. Fiind dat sistemul $\mathcal{S}$, cuplul $[\mathfrak{D}, Q_{\mathfrak{D}}[\mathcal{S}]]$ se numește structură a lui $\mathcal{S}$. În sensul acestui concept vom spune că două sisteme sînt identice dacă au aceeași structură. Să presupunem că $\mathcal{S}$ este un sistem aleator. Fie acum $\mathcal{A}$ mulțimea tuturor variablelor aleatoare, $\mathcal{X} = \{X_i / X_j \in \mathcal{A}_{i\in I}$ o parte a lui $\mathcal{A}$ și $\varphi : \mathcal{S} \to \mathcal{X}$ o aplicație care face ca fiecărui bloc $B \in \mathcal{S}$ să-i corespundă o singură variabilă $X \in \mathcal{X}$ și reciproc. Se poate arăta că $\mathcal{S}$ și $\mathcal{X}$ sînt spații topologice compacte homeomorfe în sensul clasic. Homeomorfismul φ dintre $\mathcal{S}$ și $\mathcal{X}$ face ca în locul lui $\mathcal{S}$ să fie studiat $\mathcal{X}$, iar rezultatele obținute să constituie de fapt informații asupra lui $\mathcal{S}$. Ca urmare, spunînd că un bloc B este descris în evoluția sa de o variantă aleatoare X, putem scrie $H(B)$ înțelegînd că de fapt am calculat entropia lui X. În mod analog vom scrie $H(\mathcal{T})$ și $H(\mathcal{S})$ înțelegînd că am calculat entropia vectorilor aleatori ce descriu evoluția subsistemului $\mathcal{T}$ sau a sistemului $\mathcal{S}$, acestea însă reprezentînd cantitatea de informație pe care o conține sau o furnizează sistemele $\mathcal{T}$ și respectiv $\mathcal{S}$.

Scanned with OKEN Scanner

Interpretări similare se pot da și în cazul altor indicatori numerici probabilistici corespunzători sistemului $\mathcal{S}$ sau părților acestuia. Spre exemplu, dacă blocurile B_i, $i \in I$ reprezintă caracteristici aleatoare ale unei anumite colectivități statistice, $\mathcal{S}$ va reprezenta mulțimea acestor caracteristici sau vectorul caracteristicilor considerate, iar $\mathcal{D}[\mathcal{S}]$ o grupare a acestora. Cum în astfel de situații practice caracteristicile aleatoare B_i sînt asimilate cu variabile aleatoare discrete sau continui, corespondența dintre $\mathcal{S}$ și $\mathcal{X}$ este evidentă. Indicatorii numerici ce pot fi atașați unei variabile aleatoare sau unui vector aleator (valori medii, indici de dependență etc.) pot fi gîndiți și interpretați, dacă acest lucru este necesar și interesant, pentru elementele lui $\mathcal{S}$. Aceleași considerații pot fi făcute în cazul în care blocurile B_i, $i \in I$, reprezintă factori aleatori care influențează evoluția unui anumit proces economico-social și care pot fi de asemenea gîndiți ca variabile aleatoare discrete sau continui. Numeroase alte exemple, dintre cele mai diverse, pot ilustra conceptele de sistem, subsistem, desfacere a unui sistem, precum și celelalte. Dintre acestea, o clasă importantă de sisteme o constituie sistemele cibernetice, de o natură foarte diversă, dar cu o caracteristică definitorie comună și anume aceea de conexiune inversă, sau retroacțiune, sau încă feed-back. Cu titlul de exemplificare a unor concepte ce vor fi prezentate pentru orice sistem aleator, vom face unele referiri la două sisteme cibernetice foarte diferite între ele în general și anume la sistemul de transmitere a informației și la sistemele de strategii aleatoare.

8.1.1. Sisteme de transmitere a informației

Să presupunem că sistemul $\mathcal{S}$ are desfacerea $\mathcal{D}[\mathcal{S}] = \{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3\}$, $\mathcal{T}_j \cap \mathcal{T}_k = \Phi$, $j \neq k$, $\mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}_3 = \mathcal{S}$, în care

a) $\mathcal{T}_1$ reprezintă o sursă purtătoare a unui anumit mesaj sau pe scurt o sursă de informație;

b) $\mathcal{T}_2$ reprezintă un canal de comunicație sau de transmitere a informației;

c) $\mathcal{T}_3$ reprezintă recepția semnalelor emise de $\mathcal{T}_1$ și transmise prin $\mathcal{T}_2$.

Cu aceste semnificații $\mathcal{S}$ constituie un sistem de transmitere a informației, simplu sau complex, după cum subsistemele sale sînt formate din cîte un bloc fiecare, sau cel puțin din cîte două blocuri măcar unul dintre ele. Putem vorbi

Scanned with OKEN Scanner

deci de sisteme cu una sau mai multe surse, cu unul sau mai multe canale și cu unul sau mai mulți receptori, cu o funcționare în paralel, în serie sau mixtă.

După modul în care se face emisia, transmisia și recepția informațiilor, putem vorbi de sisteme deterministe sau aleatoare. În teoria matematică a informației procesul de emisie-transmisie-recepție sau pe scurt de transmitere a informației este considerat ca un proces aleator în sensul clasic cel mai larg al acestui concept.

Semnalele emise de o sursă de informație, după o anumită lege probabilistică, pătrund pe canalul de comunicație care leagă emisia de recepție și care, într-o descriere simplă, reprezintă mediul prin care circulă semnalele purtătoare de informație. Pe canal pot să existe sau nu perturbații, prin care înțelegem tot ceea ce alterează propagarea semnalelor prin canal, făcînd ca semnalele intrate în canal să nu ajungă întotdeauna la recepție în modul în care au fost emise. Ca urmare, se poate întîmpla ca o parte din cantitatea de informație emisă să se piardă pe canal, adeseori, cînd perturbația este suficient de mare, la recepție ajungînd altceva decît ce s-a emis. Din acest motiv întrebarea „cu ce probabilitate se recepționează un anumit semnal cînd știm ce semnal se emite?" apare ca naturală pentru orice canal de comunicație, iar cunoașterea răspunsului acesteia înseamnă a cunoaște legea probabilistică a transmiterii informației prin canal.

Fie A o mulțime finită conținînd a elemente pe care le numim simboluri sau semnale elementare și cu care efectuăm transmiterea informației. În acest context, A este numită alfabet al sursei de emisie. O transmisie posibilă a sursei o reprezintă șirul

$$\omega = (\ldots, x_{-2}, x_{-1}, x_0, x_1, x_2, \ldots)\ x_i \in A,\ i \in \mathbf{Z} \qquad (8.1.2)$$

Fie $A^{\mathbf{Z}}$ mulțimea tuturor șirurilor de forma (8.1.2) și să notăm cu

$$[\alpha_1, \alpha_2, \ldots, \alpha_n;\ t_1, t_2, \ldots, t_n];$$

$$x_{t_k} = \alpha_k,\ \alpha_k \in A,\ t_k \in \mathbf{Z},\ 1 \leqslant k \leqslant n \qquad (8.1.3)$$

o parte a lui $A^{\mathbf{Z}}$ ale cărei elemente, de forma (8.1.3), se numesc cilindri. În cazul particular $t_k = k$, $1 \leqslant k \leqslant n$, un astfel de cilindru se scrie simplu sub forma $[\alpha_1, \ldots, \alpha_n]$ și se numește semnal n-dimensional.

Vom numi sursă de informație cîmpul de probabilitate $\{A^{\mathbf{Z}}, \mathcal{K}_A, \mu\}$ unde $\mathcal{K}_A$ este σ-algebra generată de mulțimea

cilindrilor (8.1.3) iar μ este o probabilitate definită pe $\mathfrak{X}_A$. Vom nota pe scurt acest cîmp cu $\{A, \mu\}$, el reprezentînd sau definind subsistemul $\mathcal{T}_1$ amintit mai sus. Probabilitatea $\mu([\alpha_1, \ldots, \alpha_n; t_1, \ldots, t_n])$ înseamnă „probabilitatea ca sursa să emită semnalul α_1 la momentul t_1, ... și semnalul, α_n la momentul t_n". Este vorba deci de transmiterea informației prin semnale n-dimensionale. Fie $T : A^Z \to A^Z$ un operator de translație temporală adică, $T\omega = \omega'$ pentru orice $\omega \in A^Z$ de forma (8.1.2) și $\omega' = (x'_k)_{k \in Z}$, $x'_k = x_{k+1}$, $k \in \mathbf{Z}$. Se spune că sursa de informație $\mathcal{T}_1 = \{A, \mu\}$ este staționară în raport cu operatorul de translație temporală T dacă pentru orice semnal compus $S \in \mathfrak{X}_A$ avem $\mu(TS) = \mu(S)$. Cu alte cuvinte, staționaritatea sursei înseamnă invariabilitatea, în raport cu timpul, a caracterului aleator al emisiei acesteia. Dacă $[\alpha_1, \ldots, \alpha_n]$ este un semnal n-dimensional în sensul celor de mai sus, mulțimea acestora constituie un cîmp de probabilitate cu a^n evenimente elementare. Cantitatea medie de informație pe care o conține acest cîmp de probabilitate este

$$H_n = -\sum_{\alpha} \mu([\alpha_1, \ldots, \alpha_n]) \log \mu([\alpha_1, \ldots, \alpha_n]) \quad (8.1.4)$$

suma efectuîndu-se după toate cele a^n semnale n-dimensionale $\alpha = [\alpha_1, \ldots, \alpha_n]$. Dacă sursa este staționară, atunci $\mu([\alpha_1, \ldots \alpha_u])$ și deci și H_n nu depind de momentul inițial, fiind în mod unic determinate de numărul n de semnale ce constituie semnalul n-dimensional. Dacă H_n constituie cantitatea medie de informație pe care o conține un semnal n-dimensional, atunci unui semnal elementar al alfabetului îi revine o cantitate medie de informație de n ori mai mică.

Definiția 8.1.1. Se numește entropie a sursei de informație oarecare $\mathcal{T}_1 = \{A, \mu\}$ și o vom nota cu $H(\mathcal{T}_1)$ cantitatea

$$H(\mathcal{T}_1) = \lim_{n \to \infty} \frac{H_n}{n} \quad (8.1.5)$$

dacă limita există.

Se poate demonstra (S. Guiașu (1968, 1977)) că orice sursă de informație staționară admite o entropie, deci că pentru orice sursă de informație staționară poate fi determinată cantitatea medie de informație pe care o conține un semnal elementar emis de sursă.

Fiind dată sursa de informație $\mathcal{T}_1 = \{A, \mu\}$ să presupunem că sînt cunoscute repartițiile probabilistice $p(x) \geqslant 0$,

Scanned with OKEN Scanner

$\sum_{x \in A} p(x) = 1$ și $p(x'/x) \geqslant 0$, $\sum_{x' \in A} p(x'/x) = 1$, $\sum_{x \in A} p(x)\, p(x'/x) = p(x')$ pentru orice $x' \in A$.

Definiția 8.1.2. Sursa de informație $\mathcal{T}_1 = \{A, \mu\}$ este de tip Markov dacă

$$\mu([\alpha_1, \ldots, \alpha_n]) = p(\alpha_1)\, p(\alpha_2/\alpha_1), \ldots, p(\alpha_n/\alpha_{n-1}) \qquad (8.1.6)$$

pentru orice semnal n-dimensional $[\alpha_1, \ldots, \alpha_n] \in \mathcal{K}_A$.

Propoziția 8.1.1. Dacă sursa de informație $\mathcal{T}_1 = \{A, \mu\}$ este de tip Markov, atunci

$$H(\mathcal{T}_1) = -\sum_{x \in A} p(x) \sum_{x' \in A} p(x'/x) \log p(x'/x) \qquad (8.1.7)$$

Demonstrație. Din (8.1.4) și (8.1.6) rezultă că

$$H_n = -\sum_{x_1, \ldots, x_n \in A} p(x_1)\, p(x_2/x_1) \ldots p(x_n/x_{n-1}) \times$$

$$\times \left[\log p(x_1) + \sum_{k=2}^{n} \log p(x_k/x_{k-1})\right] = -\sum_{x_1 \in A} p(x_1) \log p(x_1) -$$

$$-\sum_{x_1, x_2 \in A} p(x_1)\, p(x_2/x_1) \log p(x_2/x_1) -$$

$$-\sum_{x_2, x_3 \in A} \left(\sum_{x_1 \in A} p(x_1)\, p(x_2/x_1)\right) p(x_3/x_2) \log p(x_3/x_2) - \ldots -$$

$$-\sum_{x_{n-1}, x_n \in A} \left(\sum_{x_1, x_2, \ldots, x_{n-2} \in A} p(x_1)\, p(x_2/x_1) \ldots p(x_{n-1}/x_{n-2})\right) \times$$

$$\times p(x_n/x_{n-1}) \log p(x_n/x_{n-1}) = -\sum_{x \in A} p(x) \log p(x) -$$

$$-(n-1) \sum_{x \in A} p(x) \sum_{x' \in A} p(x'/x) \log p(x'/x)$$

unde $x = x_1$, care poate fi oricare din elementele lui A iar x' oricare din celelalte $(n-1)$ semnale elementare ale semnalului n-dimensional care începe cu x. Ținînd cont și de (8.1.5) rezultă că

$$H(\mathcal{T}_1) = \lim_{n \to \infty} \frac{H_n}{n} = -\sum_{x \in A} p(x) \sum_{x' \in A} p(x'/x) \log p(x'/x)$$

ceea ce demonstrează (8.1.7).

Scanned with OKEN Scanner

Să remarcăm că o sursă de informație de tip Markov este complet definită de un alfabet A, de o repartiție aleatoare definită pe A și de o matrice de tranziție stochastică definită pe același alfabet.

Definiția 8.1.3. Sursa de informație $\mathcal{T}_1 = \{A, \mu\}$ este de tip Bernoulli dacă

$$\mu([\alpha_1, \ldots, \alpha_n]) = \prod_{k=1}^{n} p(\alpha_k) \tag{8.1.8}$$

Să remarcăm că sursa de informație Bernoulli este un caz particular de sursă Markov și ca urmare, din (8.1.6) — (8.1.8) rezultă

Propoziția 8.1.2. Dacă $\mathcal{T}_1 = \{A, \mu\}$ este o sursă de informație Bernoulli, atunci

$$H(\mathcal{T}_1) = -\sum_{x \in A} p(x) \log p(x) \tag{8.1.9}$$

Să observăm că în plus, dacă semnalele alfabetului sursei Bernoulli sînt egal probabile atunci entropia sursei este

$$H(\mathcal{T}_1) = \log_2 a \tag{8.1.10}$$

Relațiile (8.1.9) și (8.1.10) constituie două exemple de regăsire a măsurilor Shannon și Hartley ale cantității de informație definite de relațiile (3.1.2) și respectiv (3.1.6).

Fie acum B o altă mulțime finită oarecare conținînd b elemente, diferită sau nu de A, mulțime care stă la baza constituirii tuturor semnalelor recepționate de la sursa de emisie $\mathcal{T}_1 = \{A, \mu\}$. Un semnal recepționat, sub forma sa cea mai generală, poate fi

$$\hat{\omega} = (\ldots, y_{-2}, y_{-1}, y_0, y_1, y_2, \ldots)\ y_k \in B,\ k \in \mathbf{Z} \tag{8.1.11}$$

Considerații asemănătoare emisiei ne conduc la mulțimea $B^{\mathbf{Z}}$ a elementelor de forma (8.1.11) și la cîmpul de probabilitate $\{B^{\mathbf{Z}}, \mathcal{K}_B, \nu_B\}$ unde $\mathcal{K}_B$ este σ-algebra generată de cilindri de forma

$$[\beta_1, \ldots, \beta_n\,;\, t_1, t_2, \ldots, t_n],\ \beta_k = y_{t_k} \in B,\ t_k \in \mathbf{Z},\ 1 \leqslant k \leqslant n, \tag{8.1.12}$$

iar probabilitatea condiționată $\nu_\omega(\hat{\omega})$ înseamnă probabilitatea de a recepționa semnalul $\hat{\omega}$ cînd s-a emis semnalul ω. Probabilitatea $\nu_\omega(\hat{\omega})$ caracterizează transmisia prin canal,

aşa cum afirmam mai sus. Într-adevăr, dacă semnalului emis ω îi corespunde un semnal $\hat{\omega}_0$, să remarcăm că dacă

$$1)\quad \nu_\omega(\hat{\omega}) = \begin{cases} 1 \text{ dacă } \hat{\omega} = \hat{\omega}_1 \neq \hat{\omega}_0 & \text{atunci perturbaţia în canal este totală, recepţionînd cu certitudine altceva decît ceea ce s-a emis} \\ 0 \text{ dacă } \hat{\omega} \neq \hat{\omega}_1 & \end{cases}$$

$$2)\quad \nu_\omega(\hat{\omega}) = \begin{cases} 1 \text{ dacă } \hat{\omega} = \hat{\omega}_0 & \text{atunci perturbaţia în canal este abesentă, recepţionînd cu certitudine ceea ce s-a emis} \\ 0 \text{ dacă } \hat{\omega} \neq \hat{\omega}_0 & \end{cases}$$

Evident că în practică transmisia se situează în general între cele două extreme 1) şi 2), avînd în general $0 < \nu_\omega(\hat{\omega}) < 1$ pentru orice cuplu de semnale, emise şi recepţionate, $(\omega, \hat{\omega}) \in A^Z \times B^Z$ şi cu cît aceste probabilităţi sînt mai mici, cu atît perturbaţia este mai mică şi deci recepţia este mai bună. Cum cantitatea de informaţie obţinută prin recepţionarea semnalului $\hat{\omega}$ cînd s-a emis semnalul ω este egală cu

$$I_\omega(\hat{\omega}) = -\log \nu_\omega(\hat{\omega}) \tag{8.1.13}$$

este evident că $I_\omega(\hat{\omega})$ este cu atît mai mare cu cît perturbaţia este mai mică, ceea ce apare ca natural şi constituie un argument în plus pentru justeţea considerării entropiei ca măsură a informaţiei.

Vom numi canal de informaţie tripletul $\{A, \nu_\omega, B\}$, acesta reprezentînd subsistemul $\mathcal{T}_2$ amintit mai sus. Vom spune că un canal $\mathcal{T}_2 = \{A, \nu_\omega, B\}$ este staţionar dacă pentru orice $\omega \in A^Z$ şi orice $S \in \mathcal{K}_B$ avem $\nu_{T\omega}(TS) = \nu_\omega(S)$, unde T este operatorul de translaţie temporală. Considerînd sursa de emisie $\mathcal{T}_1 = \{A, \mu\}$ şi canalul de comunicaţie $\mathcal{T}_2 = \{A, \nu_\omega, B\}$, să definim produsul cartezian $C = A \times B$ şi măsura λ,

$$\lambda(E \times F) = \int_E \nu_\omega(F)\, d\mu(\omega) \tag{8.1.14}$$

pentru orice $E \in \mathcal{K}_A$ şi $F \in \mathcal{K}_B$, care, ţinînd cont de măsurile μ şi ν, constituie o probabilitate a evenimentelor $E \times F$. Putem vorbi astfel de cîmpul de probabilitate $\{C^Z, \mathcal{K}_C, \lambda\}$ unde $C^Z = A^Z \times B^Z$ şi prin urmare, vom numi sursă compusă, creată prin asocierea canalului $\mathcal{T}_2 = \{A, \nu_\omega, B\}$ cu sursa de emisie $\mathcal{T}_1 = \{A, \mu\}$, sursa $\{C, \lambda\}$. Ca urmare a relaţiei (8.1.14),

pentru orice $F \in \mathfrak{K}_B$ putem defini probabilitatea $\eta(F) = \lambda(A^Z \times F)$ şi în consecinţă putem cunoaşte sursa de recepţie sau de ieşire a sursei compuse $\{B, \eta\}$, care reprezintă cel de-al treilea subsistem $\mathcal{T}_3$. Aşadar putem spune că sursa compusă reprezintă chiar sistemul de comunicaţie $\mathcal{S}$. Cunoaşterea acestuia înseamnă cunoaşterea repartiţiei probabilistice comune a emisiei şi recepţiei, de unde ar rezulta repartiţiile individuale corespunzătoare emisiei şi recepţiei precum şi repartiţia condiţionată descriind funcţionarea canalului sau invers, cunoaşterea repartiţiei emisiei şi a repartiţiei condiţionate de treceri prin canal, de unde ar rezulta repartiţia comună a emisiei şi recepţiei precum şi repartiţia recepţiei. Într-o astfel de gîndire, sistemul de comunicaţie apare ca fiind format din două subsisteme, emisia $\mathcal{T}_1$ şi recepţia $\mathcal{T}_3$, o repartiţie probabilistică individuală a canalului neprezentînd de altfel interes. Ne interesează să ştim ce trece prin canal cînd ştim ce s-a emis, sau ce trebuie să intre în canal cînd ştim ce trebuie recepţionat, ceea ce justifică necesitatea repartiţiilor condiţionate. Nu prezintă însă interes să ştim ce trece prin canal dacă nu ne interesează ce s-a emis sau ce trebuie recepţionat.

Se poate demonstra (S. Guiaşu (1968, 1977)) că dacă sursa de emisie $\mathcal{T}_1 = \{A, \mu\}$ şi canalul de comunicaţie $\mathcal{T}_2 = \{A, \nu_\omega, B\}$ sînt staţionare, atunci sursa compusă este staţionară şi ca urmare şi sursa de recepţie $\mathcal{T}_3 = \{B, \eta\}$ este staţionară. Cum orice sursă staţionară are entropie rezultă că sistemul de comunicaţie staţionar are entropie şi deci există $H(\mathcal{S}) = H(\mathcal{T}_1, \mathcal{T}_3)$, $H(\mathcal{T}_1)$ şi $H(\mathcal{T}_3)$ caracterizînd sursa compusă, sursa de intrare şi sursa de ieşire precum şi entropiile condiţionate $H(\mathcal{T}_1/\mathcal{T}_3)$ şi $H(\mathcal{T}_3/\mathcal{T}_1)$ caracterizînd canalul de comunicaţie.

După modul în care recepţia este condiţionată de emisie, canalele de comunicaţie sînt de mai multe tipuri, fiecăruia corespunzîndu-i o anumiă modalitate de determinare a repartiţiei condiţionate $\nu_\omega(\hat{\omega})$. Un astfel de tip îl constituie sistemele staţionare. Dacă A şi B sînt finite şi $\{p(y/x)\}$ este o matrice stochastică, $p(y/x) \geq 0$, $\sum_{y \in B} p(y/x) = 1$ pentru orice $x \in A$ şi $y \in B$ vom spune că $\{A, \nu_\omega, B\}$ este un canal de comunicaţie staţionar fără memorie dacă pentru orice $\omega \in A^Z$ şi orice semnal n-dimensional $[y_1, \ldots, y_n] \in \mathfrak{K}_B$ avem

$$\nu_\omega([y_1, \ldots, y_n]) = p(y_1/x_1)\, p(y_2/x_2) \ldots p(y_n/x_n). \quad (8.1.15)$$

Deducem imediat

Propoziția 8.1.3. Dacă $\mathcal{T}_2 = \{A, \nu_\omega, B\}$ este un canal staționar fără memorie atunci

$$H(\mathcal{T}_3/\mathcal{T}_1) = -\sum_{x\in A}\sum_{y\in B} p(x)\,p(y/x)\log p(y/x). \quad (8.1.16)$$

Propoziția 8.1.4. Dacă $\mathcal{T}_1 = \{A, \mu\}$ este de tip Bernoulli, iar $\mathcal{T}_2 = \{A, \nu_\omega, B\}$ este staționar fără memorie atunci $\mathcal{T}_3 = \{B, \eta\}$ este de asemenea de tip Bernoulli și

$$H(\mathcal{T}_3) = -\sum_{y\in B}\sum_{x\in A} p(x)\,p(y/x)\log[\sum_{x\in A} p(x)p(y/x)]. \quad (8.1.17)$$

Demonstrație. Ținînd cont de (8.1.4) deducem că

$$\lambda([x_1, y_1), \ldots, (x_n, y_n)]) = \prod_{k=1}^{n} p(x_k)\,p(y_k/x_k) = \prod_{k=1}^{n} p(x_k, y_k)$$

și cum

$$\eta([y_1, \ldots, y_n]) = \lambda(A^{\mathbf{Z}}\times[y_1, \ldots, y_n]) = \sum_{x_1, \ldots, x_n\in A}\prod_{k=1}^{n} p(x_k, y_k)$$

rezultă că

$$\eta([y_1, \ldots, y_n]) = \prod_{k=1}^{n} p(y_k)$$

Deci sursa de ieșire este de tip Bernoulli și conform relației (8.1.9) rezultă și relația (8.1.17) ținînd cont că

$$p(y) = \sum_{x\in A} p(x, y) = \sum_{x\in A} p(x)\,p(y/x).$$

Dacă pentru un canal de comunicație staționar fără memorie matricea stochastică de tranziție $\{p(y/x)\}$ are pe fiecare rînd aceeași repartiție probabilistică (eventual nu în aceeași ordine) $p_k \geqslant 0$, $\sum_{k=1}^{b} p_k = 1$, atunci se spune că avem un canal staționar fără memorie simetric.

8.1.2. Sisteme de strategii aleatoare

Să presupunem că avem un joc arbitrar cu n parteneri, concept care, în sensul său cel mai larg din teoria matematică a jocurilor, constituie, așa cum subliniază S. Guiașu (1973) „unul din primele exemple netriviale de sisteme cu conexiune inversă". Cu toate că acest concept poate fi explicat în diferite moduri, în esență jocul este compus dintr-o mulțime de mutări

Scanned with OKEN Scanner

sau acţiuni succesive efectuate de către parteneri, după o regulă apriori stabilită, ale căror interese pot fi aceleași sau diferite, uneori chiar contradictorii. Orice acţiune a unuia sau unora dintre parteneri depinde de modul în care au acționat ceilalţi parteneri și la rîndul ei influenţează acţiunile viitoare ale partenerilor de joc. Să notăm cu $J_1, J_2, \ldots, J_n$ cei n parteneri sau jucători și cu $\mathcal{S}$ mulţimea acestora, reprezentanţi fizici ai unor sisteme complete de probabilităţi atașate unor mulţimi bine definite de mutări sau acţiuni ale partenerilor de joc. Jucătorul J_i poate efectua la un moment dat mutarea $m_i^{k_i}$, $1 \leqslant k_i \leqslant n_i$, avînd la dispoziţie n_i astfel de mutări, cu probabilitatea $p(m_i^{k_i})$. Strategia acestuia, reprezentînd deci variabila aleatoare ale cărei valori sînt mutările $m_i^{k_i}$, o vom nota cu

$$X_i : \begin{pmatrix} m_i^{k_i} \\ \\ p(m_i^{k_i}) \end{pmatrix}; \; p(m_i^{k_i}) \geqslant 0, \; \sum_{k_i=1}^{n_i} p(m_i^{k_i}) = 1, \; 1 \leqslant i \leqslant n \qquad (8.1.18)$$

și înţelegem prin aceasta mulţimea mutărilor pe care jucătorul la are la dispoziţie împreună cu probabilităţile de a le utiliza. O mutare efectuată la un moment dat, succesiv, de cei n jucători, poate fi reprezentată de vectorul $(m_1^{k_1}, m_2^{k_2}, \ldots, m_i^{k_i}, \ldots, m_n^{k_n})$ cu probabilitatea $p(m_1^{k_1}, m_2^{k_2}, \ldots, m_i^{k_i}, \ldots, m_n^{k_n})$, $1 \leqslant k_i \leqslant n_i$, $1 \leqslant i \leqslant n$. Să notăm cu $\mathcal{S} = \{J_i\}_{i=1}^n$ și cu $\mathfrak{X} = \{X_i\}_{i=1}^n$ cele două sisteme homeomorfe, fiecărui jucător J_i corespunzîndu-i pe parcursul unui joc o singură strategie aleatoare X_i și reciproc. În acest context, orice desfacere $\mathfrak{D}[\mathcal{S}]$ de forma (8.1.1) poartă numele de coalizare a jucătorilor, fiecare subsistem notat acum cu $\mathcal{C}$ numindu-se coaliţie. O coaliţie $\mathcal{C}$ este trivială dacă aceasta conţine un singur jucător sau dacă aceasta cuprinde toţi jucătorii. Comportamentul unui jucător, ca mod de acţionare, pe parcursul desfășurării jocului fiind descris de strategia adoptată, deci de o variabilă aleatoare, rezultă că modul de acţiune sau comportamentul în joc al unei coaliţii va fi descris de un vector aleator. Cunoașterea strategiilor (8.1.18) implică posibilitatea de determinare a cantităţii de informaţie sau a entropiei sistemului $\mathcal{S}$, $H(\mathcal{S})$, a unei coaliţii $\mathcal{C}$, $H(\mathcal{C})$ sau a unui jucător J, $H(J)$ precum și diverse entropii condiţionate $H(\mathcal{C}_j/\mathcal{C}_k)$ reprezentînd cantitatea medie de informaţie proprie coaliţiei $\mathcal{C}_j$ cînd se cunoaște comportamentul coaliţiei $\mathcal{C}_k$, entropii ale căror proprietăţi sînt date în paragraful 3.3.

Scanned with OKEN Scanner

Ca urmare putem scrie că $H(\mathcal{C}_j/\mathcal{C}_k) = H(\mathcal{C}_j, \mathcal{C}_k) - H(\mathcal{C}_k)$, $H(\mathcal{C}_j/\mathcal{C}_k) \leqslant H(\mathcal{C}_j)$, $1 \leqslant j \neq k \leqslant m$ și că $H(J) \leqslant H(\mathcal{C}) \leqslant H(\mathcal{S})$ pentru orice $J \subset \mathcal{C} \subset \mathcal{S}$, ceea ce arată că în cazul unui joc cu strategii aleatoare gradul de nedeterminare este monoton crescător în raport cu numărul partenerilor. Vom spune că două coaliții $\mathcal{C}_j$ și $\mathcal{C}_k$ sînt entropic independente pe parcursul unui joc dacă $H(\mathcal{C}_j/\mathcal{C}_k) = H(\mathcal{C}_j)$. De asemenea, vom spune că o coaliție $\mathcal{C}_j$ este complet dependentă pe parcursul unui joc de coaliția $\mathcal{C}_k$ dacă $H(\mathcal{C}_j/\mathcal{C}_k) = 0$ și că acestea sînt reciproc complet dependente dacă $H(\mathcal{C}_j/\mathcal{C}_k) = H(\mathcal{C}_k/\mathcal{C}_j) = 0$, $1 \leqslant j \neq k \leqslant m$.

Dacă mutării $m_i^{k_i}$ îi asociem și o utilitate $u(m_i^{k_i}) = u_i^{k_i}$, atunci am putea considera drept cantitate medie de informație furnizată de jucătorul J_i prin adoptarea strategiei X_i, $1 \leqslant i \leqslant n$, mărimea

$$\Gamma_{n_i}[u_i ; J_i] = - \sum_{k_i}^{n_i} u(m_i^{k_i})\, p(m_i^{k_i}) \log p(m_i^{k_i}), \quad (8.1.19)$$

adică tocmai entropia ponderată (4.2.1), mărime care devine maximă în condițiile propoziției 4.2.4, a cărei formulare ar deveni acum

Propoziția 8.1.5. Cantitatea medie de informație ponderată (8.1.19) a jucătorului J_i devine maximă dacă și numai dacă strategia aleatoare X_i adoptată de acesta urmează repartiția

$$p(m_i^{k_i}) = \frac{1}{\mathrm{e}} 2^{\frac{\alpha}{u(m_i^{k_i})}}, \quad 1 \leqslant k_i \leqslant n_i \quad (8.1.20)$$

unde α este soluția ecuației

$$\sum_{k_i=1}^{n_i} 2^{\frac{\alpha}{u(m_i^{k_i})}} = \mathrm{e}, \quad (8.1.21)$$

caz în care

$$[\max_{x_i}]\, \Gamma_{n_i}[u_i ; J_i] = - \frac{\alpha}{\mathrm{e}} + \frac{\log \mathrm{e}}{\mathrm{e}} \sum_{k_i=1}^{n_i} (u(m_i^{k_i})\, 2^{\frac{\alpha}{u(m_i^{k_i})}}. \quad (8.1.22)$$

Într-un sens mai restrîns, utilitatea unei mutări poate fi interpretată drept cîștig ce ar reveni jucătorului în chestiune. În acest caz utilitatea poate fi și negativă, o anumită

Scanned with OKEN Scanner

mutare putînd aduce jucătorului o pierdere, deci un cîştig negativ. Indicatorul care se analizează în astfel de situaţii este acela de utilitate medie sau cîştig mediu pe parcursul unui joc sau partide. În anumite jocuri însă, se poate întîmpla ca utilitatea fiecărei mutări a fiecărui jucător să fie nenegativă. Dacă $u^i(m_1^{k_1}, \ldots, m_n^{k_n})$ este utilitatea individuală corespunzătoare unei mutări n-dimensionale efectuată pe rînd de toţi partenerii de joc la un moment dat, putînd avea $n_1 \times n_2 \times \ldots \times n_n$ astfel de mutări, atunci, prin definiţie, utilitatea medie corespunzătoare jucătorului J_i, $1 \leqslant i \leqslant n$, este dată de relaţia

$$U(J_i) = \sum_{k_1=1}^{n_1} \ldots \sum_{k_n=1}^{n_n} u^i(m_1^{k_1}, \ldots, m_n^{k_n})\, p(m_1^{k_1}) \ldots p(m_n^{k_n}). \quad (8.1.23)$$

Dar, odată cu utilitatea medie totală a jocului din punctul de vedere al jucătorului J_i, $1 \leqslant i \leqslant n$, aflat într-o anumită conexiune cu ceilalţi parteneri, adoptarea unei anumite strategii aleatoare conduce şi la înlăturarea unei anumite nedeterminări sau la obţinerea unei anumite cantităţi de informaţie. Ţinînd cont de cele două concepte, de altfel distincte, de utilitate şi de informaţie, se poate defini beneficiul total mediu mixt al jucătorului J_i, $1 \leqslant i \leqslant n$, ca fiind

$$\mathfrak{B}_i(J_1, \ldots, J_n) = U(J_i) + H(J_i), \quad 1 \leqslant i \leqslant n. \quad (8.1.24)$$

Soluţionînd o problemă de extremum cu legături, deducem cu uşurinţă

Propoziţia 8.1.6. Strategia optimă a jucătorului J_i pentru care beneficiul său total mediu mixt devine maxim, este dată de relaţia

$$p(m_i^{k_i}) = \frac{1}{e}\, 2^{\alpha + A(U(J_i))}, \quad 1 \leqslant k_i \leqslant n_i, \quad 1 \leqslant i \leqslant n \quad (8.1.25)$$

unde

$$A(U(J_i)) = \sum_{\substack{s=1 \\ s \neq i}}^{n} \sum_{k_s=1}^{n_s} \left[u^i(m_1^{k_1}, \ldots, m_n^{k_n}) \prod_{\substack{j=1 \\ j \neq i}}^{n} p(m_j^{k_j}) \right], \quad (8.1.26)$$

iar α este soluţie a ecuaţiei

$$\sum_{k_i=1}^{n_i} 2^{A(U(J_i))} = e\, 2^{-\alpha}. \quad (8.1.27)$$

Scanned with OKEN Scanner

Din relaţiile (8.1.24) şi (8.1.25) putem scrie şi expresia beneficiului maxim $\mathfrak{B}_i^{\max}(J_1, \ldots, J_n)$ pe care însă nu o dăm fiind greoaie în cazul general. Ca o consecinţă imediată a propoziţiei 8.1.6 rezultă

Propoziţia 8.1.7. Dacă

$$u^i(m_1^{k_1}, \ldots, m_n^{k_n}) = u^i(m_i^{k_i}) = u_{k_i}^i, \qquad 1 \leqslant k_i \leqslant n_i, \quad 1 \leqslant i \leqslant n \tag{8.1.28}$$

atunci strategia optimă a jucătorului J_i este definită de relaţia

$$p(m_i^{k_i}) = \frac{1}{\mathrm{e}} 2^{\alpha + u_{k_i}^i}, \quad 1 \leqslant k_i \leqslant n_i, \quad 1 \leqslant i \leqslant n, \tag{8.1.29}$$

unde α este soluţie a ecuaţiei

$$\sum_{k_i=1}^{n_i} 2^{u_{k_i}^i} = \mathrm{e}\, 2^{-\alpha} \tag{8.1.30}$$

caz în care

$$\mathfrak{B}_i^{\max}(J_1, \ldots, J_n) = \log \mathrm{e} - \alpha = \log \left[\sum_{k_i=1}^{n_i} 2^{u_{k_i}^i} \right]. \tag{8.1.31}$$

Rezultatul de mai sus poate fi generalizat dacă în locul entropiei Shannon $H(J_i)$ se consideră entropia ponderată $\Gamma[u^i, J_i]$, în expresia (8.1.24). În acest caz obţinem fără dificultate

$$p(m_i^{k_i}) = \frac{2}{\mathrm{e}} \cdot 2^{\frac{\alpha}{u_{k_i}^i}}, \quad 1 \leqslant k_i \leqslant n_i, \quad 1 \leqslant i \leqslant n \tag{8.1.32}$$

fiind soluţia ecuaţiei

$$\sum_{k_i=1}^{n_i} 2^{\frac{\alpha}{u_{k_i}^i}} = \frac{\mathrm{e}}{2}, \tag{8.1.33}$$

iar beneficiul maxim are expresia

$$\mathfrak{B}_i^{\max}(J_1, \ldots, J_n) = \frac{2 \log \mathrm{e}}{\mathrm{e}} \sum_{k_i=1}^{n_i} u_{k_i}^i 2^{\frac{\alpha}{u_{k_i}^i}} - \alpha. \tag{8.1.34}$$

Noţiunea de utilitate apare şi cînd trecem de la un jucător la o coaliţie şi ca urmare, rezultatele de mai sus pot fi gene-

ralizate în acest sens pentru o coaliție $\mathcal{C}_j$, $1 \leqslant j \leqslant m$, scrierea lor fiind însă mai anevoioasă. În cazul particular al unui joc cu trei parteneri, sau triadă, diverse rezultate interesante și riguros analizate, sînt expuse în lucrările lui S. Guiașu (1973, 1977), pretutindeni pornindu-se de la conceptul de cantitate de informație liberă sau ponderată.

8.2. Măsuri entropice ale interdependenței

În evoluția sistemului $\mathcal{S}$, între părțile sale, blocuri sau subsisteme, există anumite conexiuni, sau altfel spus, o anumită coeziune sau interdependență care fac ca orice modificare efectuată asupra unui element al lui $\mathcal{S}$ să se răsfrîngă sau nu asupra altor elemente ale acestuia, după cum acestea depind sau nu depind unele de altele. Apare deci ca necesară posibilitatea măsurării gradului în care unele elemente ale sistemului depind de altele, sau a gradului de interdependență dintre diferite elemente ale lui $\mathcal{S}$, ceea ce justifică necesitatea măsurilor dependenței unilaterale sau a celei globale. Dată fiind corespondența dintre sistemele $\mathcal{S}$ și $\mathcal{X}$, și cum pentru măsurarea dependenței dintre variabile aleatoare sau vectori aleatori pot fi utilizate măsurile prezentate în capitolele 6 și 7, se poate spune că problema determinării gradului de dependență probabilistă dintre elementele constitutive ale sistemului $\mathcal{S}$ este soluționată. Mai mult însă, se poate afirma, în baza rezultatelor din aceleași capitole, că o soluționare generală a acestei probleme o constituie doar corelația entropică, întrucît cu ajutorul acesteia putem măsura gradul de dependență globală sau de interdependență dintre oricîte părți ale lui $\mathcal{S}$ și de asemenea gradul în care depind una de alta orice două părți ale lui $\mathcal{S}$. Astfel, măsura (7.6.19) devine acum

$$W[\mathcal{S}; \mathfrak{D}[\mathcal{S}]] = \sum_{\mathcal{T} \in \mathfrak{D}[\mathcal{S}]} H(\mathcal{T}) - H(\mathcal{S}) \qquad (8.2.1)$$

și reprezintă gradul de dependență globală sau de interdependență dintre subsistemele lui $\mathfrak{D}[\mathcal{S}]$ a căror reuniune constituie sistemul $\mathcal{S}$. În cazul particular al desfacerii atomice $\mathfrak{D}[\mathcal{S}] = \mathcal{S}$, relația (8.2.1) devine

$$W[\mathcal{S}; B_1, \ldots, B_n] = \sum_{i=1}^{n} H(B_i) - H(\mathcal{S}) \qquad (8.2.2)$$

reprezentînd gradul de interdependenţă sau intensitatea coeziunii dintre blocurile constitutive ale sistemului $\mathcal{S}$. În baza propoziţiei 7.6.9 avem totdeauna

$$W[\mathcal{S}; B_1, \ldots, B_n] \geqslant \max_{\mathfrak{D}[\mathcal{S}]} W[\mathcal{S}; \mathfrak{D}[\mathcal{S}]] \qquad (8.2.3)$$

ceea ce înseamnă că cel mai mare grad de interdependenţă pe care îl putem găsi în diferitele moduri în care este reprezentat $\mathcal{S}$ este dat de reprezentarea atomică a sistemului. Cu alte cuvinte, în baza relaţiei (8.2.3) rezultă că interdependenţa dintre grupurile de elemenete ale lui $\mathcal{S}$ este mai mică decît interdependenţa dintre toate elementele lui $\mathcal{S}$.

În mod analog putem prezenta coeficienţii de corelaţie entropică (7.6.28) care pot fi scrişi acum sub forma

$$G_i[\mathcal{S}; \mathfrak{D}[\mathcal{S}]] = \frac{W[\mathcal{S}; \mathfrak{D}[\mathcal{S}]]}{M_i[\mathfrak{D}(\mathcal{S}]]}, \quad 1 \leqslant i \leqslant 4, \qquad (8.2.4)$$

Toate rezultatele prezentate în paragraful 7.6.1 asupra măsurilor entropice ale dependenţei globale rămîn valabile în cazul de faţă cu adaptarea corespunzătoare a terminologiei, afirmaţie valabilă de asemenea şi cînd vorbim despre măsurarea gradului în care subsistemul $\mathcal{T}_j$ depinde de subsistemul $\mathcal{T}_k$, $\mathcal{T}_j \subset \mathcal{S}$, $\mathcal{T}_k \subset \mathcal{S}$, $\mathcal{T}_j \cap \mathcal{T}_k = \emptyset$, folosind în acest scop măsura (7.6.38) a cărei formă poate fi acum

$$U(\mathcal{T}_j, \mathcal{T}_k) = \frac{H(\mathcal{T}_j) - H(\mathcal{T}_j/\mathcal{T}_k)}{H(\mathcal{T}_j)} \qquad (8.2.5)$$

şi în baza căreia constatăm, ceea ce apare de altfel ca natural, că în general gradul în care $\mathcal{T}_j$ depinde de $\mathcal{T}_k$ nu este acelaşi cu gradul în care $\mathcal{T}_k$ depinde de $\mathcal{T}_j$. Legătura între cele două tipuri de dependenţă globală şi unilaterală este asigurată de relaţia (7.6.42) care devine acum

$$G_1[\mathcal{T}_j, \mathcal{T}_k] = \alpha_{jk} U(\mathcal{T}_j, \mathcal{T}_k) + \alpha_{kj} U(\mathcal{T}_k, \mathcal{T}_j) \qquad (8.2.6)$$

şi în virtutea căreia, dependenţa globală dintre subsistemele $\mathcal{T}_j$ şi $\mathcal{T}_k$ apare ca o combinaţie liniară de dependenţele unilaterale dintre cele două subsisteme, coeficienţii α_{jk} şi α_{kj} fiind

$$\alpha_{jk} = \frac{H(\mathcal{T}_j)}{2H(\mathcal{T}_j, \mathcal{T}_k)}; \qquad \alpha_{kj} = \frac{H(\mathcal{T}_k)}{2H(\mathcal{T}_j, \mathcal{T}_k)}. \qquad (8.2.7)$$

Scanned with OKEN Scanner

Ca urmare a relaţiilor (7.6.36) şi (8.2.6) putem deduce

Propoziţia 8.2.1. Între dependenţa globală şi dependenţele unilaterale corespunzătoare subsistemelor $\mathcal{T}_1$, $\mathcal{T}_2$ şi $\mathcal{T}_3$ ale lui $\mathcal{S}$ are loc relaţia

$$G_1[\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3] = \sum_{\substack{i,j=1 \\ i<j}}^{3} [\alpha_{ij} U(\mathcal{T}_i, \mathcal{T}_j) + \alpha_{ji} U(\mathcal{T}_j, \mathcal{T}_i)] +$$

$$+ \sum_{\substack{i,j,k=1 \\ i<j \\ k\neq i, k\neq j}}^{3} [\alpha_{ij,k} U[(\mathcal{T}_i, \mathcal{T}_j), \mathcal{T}_k] + \alpha_{k,ij} U[\mathcal{T}_k, (\mathcal{T}_i, \mathcal{T}_j)]] \quad (8.2.8)$$

unde α_{ij}, α_{ji}, $\alpha_{ij,k}$ şi $\alpha_{k,ij}$ sînt funcţii de entropiile directe ale celor trei subsisteme.

Demonstraţie. Să observăm că putem scrie

$$G_1[\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3] = \frac{1}{3} \sum_{\substack{i,j=1 \\ i<j}}^{3} \frac{H(\mathcal{T}_i, \mathcal{T}_j)}{H(\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3)} G_1[\mathcal{T}_i, \mathcal{T}_j] +$$

$$+ \frac{1}{3} \sum_{\substack{i,j,k=1 \\ i<j, k\neq i, k\neq j}}^{3} G_1[(\mathcal{T}_i, \mathcal{T}_j), \mathcal{T}_k]. \quad (8.2.9)$$

Ţinînd cont de relaţia (8.2.6) şi notînd cu

$$\alpha_{ij} = \frac{1}{6} \frac{H(\mathcal{T}_i)}{H(\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3)}; \quad \alpha_{ij,k} = \frac{1}{6} \frac{H(\mathcal{T}_i, \mathcal{T}_j)}{H(\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3)};$$

$$\alpha_{k,ij} = \frac{1}{6} \frac{H(\mathcal{T}_k)}{H(\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3)} \quad (8.2.10)$$

deducem relaţia (8.2.8).

Ca urmare a propoziţiei 8.2.1 deducem că pentru o anumită desfacere a lui $\mathcal{S}$, $\mathcal{D}[\mathcal{S}] = \{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3\}$, sau dimpotrivă pentru o reconstituire a acestuia alipind la cuplul $(\mathcal{T}_i, \mathcal{T}_j)$ subsistemul $\mathcal{T}_k$, $1 \leqslant i \neq j \neq k \leqslant 3$, între interdependenţa sau coeziunea tripletului şi dependenţele unilaterale ale elementelor constitutive are loc relaţia

$$G_1[\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3] = a_{ij} U(\mathcal{T}_i, \mathcal{T}_j) + a_{ji} U(\mathcal{T}_j, \mathcal{T}_i) +$$

$$+ b_{ij,k} U[(\mathcal{T}_i, \mathcal{T}_j), \mathcal{T}_k] + b_{k,ij} U[\mathcal{T}_k, (\mathcal{T}_i, \mathcal{T}_j)] \quad (8.2.11)$$

unde $a_{ij} = 3\alpha_{ij}$, $b_{ij,k} = 3\alpha_{ij,k}$ şi $b_{k,ij} = 3\alpha_{k,ij}$.

Scanned with OKEN Scanner

În cazul în care pentru oricare din elementele sistemului $\mathcal{S}$ se are în vedere entropia ponderată sau cu alte cuvinte cantitatea medie de informație utilă, atunci, în virtutea rezultatelor din paragraful 7.6.3 putem vorbi despre măsuri entropice ponderate ale gradului de interdependență, în care relația (7.6.67) devine acum

$$W[u\,;\,\mathcal{S},\mathfrak{D}[\mathcal{S}]] = \sum_{\mathcal{T}\in\mathfrak{D}[\mathcal{S}]} \Gamma[u\,;\,\mathcal{T}] - \Gamma[u\,;\mathcal{S}] \qquad (8.2.12)$$

și extinde relația (8.2.1) pe care o putem considera măsură entropică liberă a dependenței.

Să presupunem acum că $\mathcal{S} = \{\mathcal{T}_1,\ \mathcal{T}_2,\ \mathcal{T}_3\}$ este un sistem de comunicație. Ca urmare, putem analiza interdependența dintre emisia $\mathcal{T}_1$ și recepția $\mathcal{T}_3$ în baza relației (8.2.1) care devine

$$W[\mathcal{T}_1,\ \mathcal{T}_3] = H(\mathcal{T}_1) + H(\mathcal{T}_3) - H(\mathcal{T}_1,\ \mathcal{T}_3) \qquad (8.2.13)$$

sau sub o altă formă

$$W\,[\mathcal{T}_1, \mathcal{T}_3] = H(\mathcal{T}_1) - H\,(\mathcal{T}_1/\mathcal{T}_3) =$$

$$= H(\mathcal{T}_3) - H(\mathcal{T}_3/\mathcal{T}_1) \qquad (8.2.14)$$

unde atît $\mathcal{T}_1$ cît și $\mathcal{T}_3$ pot fi blocuri sau mulțimi de blocuri, adică putem avea atît emisia cît și recepția unibloc sau multibloc, într-un sistem de comunicație general. Ca urmare a faptului că entropiile condiționate $H(\mathcal{T}_1/\mathcal{T}_3)$ și $H(\mathcal{T}_3/\mathcal{T}_1)$ caracterizează canalul de comunicație, rezultă că interdependența dintre emisie și recepție trebuie privită implicit ca fiind de fapt interdependența dintre emisie, canal și recepție și deci ca dependență globală între cele trei subsisteme ale lui $\mathcal{S}$.

Ca urmare a relației (3.3.10) am putea spune că $H(\mathcal{T}_1) - H(\mathcal{T}_1/\mathcal{T}_3)$ reprezintă cantitatea de informație pe care sursa de emisie o „datorează" cunoașterii a ceea ce trebuie recepționat, iar $H(\mathcal{T}_3) - H(\mathcal{T}_3/\mathcal{T}_1)$ este cantitatea de informație pe care recepția o „datorează" cunoașterii a ceea ce s-a emis. În baza relației (8.2.13) rezultă egalitatea celor două cantități de informație, calitativ diferite. În virtutea considerațiilor din paragraful 3.3.1 se poate spune că sursa

Scanned with OKEN Scanner

de emisie $\mathcal{T}_1$ și sursa de recepție $\mathcal{T}_3$ sînt independente sau că „nu își datorează" nimic una alteia dacă $H(\mathcal{T}_1) - H(\mathcal{T}_1/\mathcal{T}_3) = H(\mathcal{T}_3) - H(\mathcal{T}_3/\mathcal{T}_1) = 0$. De asemenea vom spune că emisia „datorează" totul recepției dacă $H(\mathcal{T}_1) - H(\mathcal{T}_1/\mathcal{T}_3) = H(\mathcal{T}_1)$, ceea ce înseamnă că $H(\mathcal{T}_1/\mathcal{T}_3) = 0$, precum și că ele își „datorează" totul una alteia sau că sînt reciproc complet dependente dacă $H(\mathcal{T}_1/\mathcal{T}_3) = H(\mathcal{T}_3/\mathcal{T}_1) = 0$. Interpretînd entropia condiționată $H(\mathcal{T}_3/\mathcal{T}_1)$ ca fiind cantitatea medie de informație care se pierde pe canal datorită perturbațiilor atunci, $H(\mathcal{T}_3) - H(\mathcal{T}_3/\mathcal{T}_1)$ reprezintă cantitatea medie de informație conținută de un semnal recepționat de sursa $\mathcal{T}_3$ și ca urmare a relației (8.2.13) aceasta este egală cu cantitatea medie de informație furnizată de un semnal emis de sursa $\mathcal{T}_1$ și recepționat de sursa $\mathcal{T}_3$. Dacă pierderea de informație se datorează canalului, atunci evident că informația medie pe care sursa de recepție o primește depinde, în baza relației (8.2.13), în mod esențial de sursa de emisie $\mathcal{T}_1 = \{A, \mu\}$, deci de repartiția μ în funcție de care entropia sursei $H(\mathcal{T}_1)$ este mai mică sau mai mare. Să notăm cu

$$V_\mu[\mathcal{T}_1, \mathcal{T}_3] = H(\mathcal{T}_1) - H(\mathcal{T}_1/\mathcal{T}_3) \qquad (8.2.15)$$

cantitatea medie de informație conținută de un semnal emis de sursa $\mathcal{T}_1 = \{A, \mu\}$ și care ajunge la recepție, expresie care se numește viteză de transmitere a informației produse de sursa $\mathcal{T}_1 = \{A, \mu\}$ și transmisă prin canalul de comunicație $\mathcal{T}_2 = \{A, \nu_\omega, B\}$.

Ca urmare a propozițiilor 3.3.5, 7.6.1 și 7.6.2 putem formula imediat

Propoziția 8.2.2. Următoarele afirmații asupra vitezei de transmitere a informației $V_\mu[\mathcal{T}_1, \mathcal{T}_3]$ sînt adevărate întotdeauna :

1) este definită pentru orice sistem de transmitere a informației cu evoluție aleatoare staționară ;

2) $V_\mu[\mathcal{T}_1, \mathcal{T}_3] = V_\eta[\mathcal{T}_3, \mathcal{T}_1]$;

3) $0 \leqslant V_\mu[\mathcal{T}_1, \mathcal{T}_3] < \infty$;

4) $V_\mu[\mathcal{T}_1, \mathcal{T}_3] = 0$ dacă și numai dacă emisia și recepția sînt independente ;

5) dacă sursa de emisie $\mathcal{T}_1$ este complet dependentă de sursa de recepție $\mathcal{T}_3$ atunci $V\mu[\mathcal{T}_1, \mathcal{T}_3] = H(\mathcal{T}_1)$ și

Scanned with OKEN Scanner

analog dacă recepția este complet dependentă de emisie;

6) dacă emisia $\mathcal{T}_1$ și recepția $\mathcal{T}_3$ sînt reciproc complet dependente, atunci $V_\mu[\mathcal{T}_1, \mathcal{T}_3] = H(\mathcal{T}_1) = H(\mathcal{T}_3) \leqslant$ $= H(\mathcal{T}_1, \mathcal{T}_3)$;

7) $V_\mu[\mathcal{T}_1, \mathcal{T}_3] \leqslant \min\{H(\mathcal{T}_1), H(\mathcal{T}_3)\} \leqslant \frac{1}{2}\{H(\mathcal{T}_1) + H(\mathcal{T}_3)\} \leqslant$ $\leqslant \max\{H(\mathcal{T}_1), H(\mathcal{T}_3)\} \leqslant H(\mathcal{T}_1, \mathcal{T}_3)$ cu egalitate peste tot dacă și numai dacă emisia și recepția sînt reciproc complet dependente;

8) $V_\mu[\mathcal{T}_1, \mathcal{T}_3]$ este invariantă în raport cu orice transformare biunivocă a procesului de transmitere a informației.

Din relațiile (8.2.14) și (8.2.15) deducem că valoric viteza de transmitere a informației prin canal și gradul de interdependență dintre emisie și recepție coincid, diferența dintre ele fiind numai calitativă. Pe de altă parte deducem cu ușurință că gradul de dependență globală dintre emisie și recepție este cu atît mai mare cu cît cantitatea medie de informație care ajunge de la emisie la recepție este mai mare.

Viteza de transmitere a informației fiind dată pentru o anumită sursă de emisie $\mathcal{T}_1 = \{A, \mu\}$, deci pentru o anumită lege probabilistă de emisie a semnalelor μ, apare natural să considerăm și viteza maximă de transmitere a informației. Fiind dată sursa de emisie $\mathcal{T}_1 = \{A, \mu\}$ și canalul de comunicație $\mathcal{T}_2 = \{A, \nu_\omega\, B\}$ se numește capacitate a canalului viteza maximă de transmitere a informației

$$K[\mathcal{T}_2] = K[\{A, \nu_\omega, B\}] = \sup_\mu \{V_\mu[\mathcal{T}_1, \mathcal{T}_3]\}. \qquad (8.2.16)$$

Să remarcăm că putem formula și pentru capacitatea unui canal o propoziție analogă propoziției 8.2.2. Capacitatea K apare în mod direct ca o caracteristică a unui canal de transmitere a informației și a da un canal cu capacitatea K înseamnă a specifica faptul că prin acest canal nu poate fi transmisă o cantitate de informație, într-un anumit scop, superioară lui K. Indirect însă, capacitatea canalului apare ca o caracteristică a întregului sistem de transmitere a informației $\mathcal{S} = \{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3\}$, unde

$$\mathcal{T}_1 = \{A, \mu\}, \quad \mathcal{T}_2 = \{A, \nu_\omega, B\} \quad \text{și} \quad \mathcal{T}_3 = \{B, \eta\}.$$

Scanned with OKEN Scanner

Ținînd cont de propozițiile 8.1.3 și 8.1.4 deducem

Propoziția 8.2.3. Într-un sistem de transmitere a informației cu emisie de tip Bernoulli și canal staționar fără memorie capacitatea canalului este

$$K[\mathcal{T}_2] = \sup_{p(x)} \left\{ \sum_{x \in A} \sum_{y \in B} p(x)\, p(y/x) \log \frac{p(y/x)}{\sum\limits_{x \in A} p(x) p(y/x)} \right\} \tag{8.2.17}$$

supremumul fiind considerat după toate repartițiile $p(x) \geqslant 0$, $\sum\limits_{x \in A} p(x) = 1$.

Se spune că un canal de transmitere a informației staționar fără memorie este simetric dacă matricea stochastică de tranziție $[p(y/x)]$ are pe fiecare linie aceeași repartiție (eventual nu în aceeași ordine) $p_k \geqslant 0$, $\sum\limits_{k=1}^{b} p_k = 1$. Ca urmare, din (8.2.17) rezultă

Propoziția 8.2.4. Într-un sistem de comunicație cu emisie de tip Bernoulli și canal staționar fără memorie și simetric capacitatea canalului este

$$K[\mathcal{T}_2] = \log b + \sum_{k=1}^{b} p_k \log p_k. \tag{8.2.18}$$

Presupunînd că fiecare semnal transmis are o anumită importanță sau utilitate, să notăm cu $u(x, y) \geqslant 0$ utilitatea perechii de semnale (x, y), x fiind semnalul emis și y semnalul recepționat. Ținînd cont de relația (7.6.52), în condițiile propoziției 8.2.3. relația (8.2.17) devine

$$K[\mathcal{T}_2] = \sup_{p(x)} \left\{ \sum_{x \in A} \sum_{y \in B} u(x, y) p(x) p(y/x) \log \frac{p(y/x)}{\sum\limits_{x \in A} p(x) p(y/x)} \right\} \tag{8.2.19}$$

și reprezintă capacitatea ponderată a unui canal staționar fără memorie și simetric, ilustrînd astfel un caz concret de folosire a măsurilor entropice ponderate ale dependenței.

Să presupunem acum că $\mathcal{S} = \{J_i\}_{i=1}^{n}$, deci că avem un joc cu n parteneri cu strategii aleatoare. În acest caz, mărimea (8.2.2) reprezintă gradul de interdependență dintre cei n

Scanned with OKEN Scanner

parteneri iar mărimea (8.2.1) reprezintă gradul de interdependenţă dintre cele m coaliţii (8.1.1) ce pot fi realizate cu aceştia. În baza relaţiei (8.2.3) rezultă că prin coalizarea jucătorilor, gradul de interdependenţă dintre elementele lui $D[\mathcal{S}]$, jucători sau coaliţii, participante la joc, este descrescător, cea mai mare coeziune la nivel de ansamblu existînd în cazul jocului individual, ceea ce trebuie să apară de altfel ca natural. Relaţiile (8.2.8) şi (8.2.11) ilustrează legătura dintre dependenţa globală a unui triplet de coaliţii sau jucători şi gradul de dependenţă unilaterală dintre aceştia.

8.3. Metrici entropice

Fie $\mathcal{S} = \{B_i\}_{i \in I}$ un sistem real, $\mathcal{D}[\mathcal{S}]$ o descompunere a sa (8.1.1) şi $d : \mathcal{D}[\mathcal{S}] \times \mathcal{D}[\mathcal{S}] \to [0, \infty)$ o funcţie care are următoarele proprietăţi :

a) $d(\mathcal{T}_j, \mathcal{T}_k) = 0$ dacă şi numai dacă $\mathcal{T}_j = \mathcal{T}_k$;

b) $d(\mathcal{T}_j, \mathcal{T}_k) = d(\mathcal{T}_k, \mathcal{T}_j)$ pentru orice $\mathcal{T}_j$ şi $\mathcal{T}_k$ din $\mathcal{D}[\mathcal{S}]$;

c) $d(\mathcal{T}_j, \mathcal{T}_k) \leqslant d(\mathcal{T}_j, \mathcal{T}_l) + d(\mathcal{T}_l, \mathcal{T}_k)$ pentru orice $\mathcal{T}_j, \mathcal{T}_k$ şi $\mathcal{T}_l$ din $\mathcal{D}[\mathcal{S}]$.

Funcţia d se numeşte metrică în $\mathcal{D}[\mathcal{S}]$ iar perechea $(\mathcal{D}[\mathcal{S}], d)$ se numeşte spaţiu metric. Să observăm că $d(\mathcal{T}_j, \mathcal{T}_k)$ reprezintă distanţa dintre două subsisteme ale lui $\mathcal{S}$, iar dacă $\mathcal{D}[\mathcal{S}] = \mathcal{S}$ atunci avem distanţa dintre două blocuri ale lui $\mathcal{S}$.

Propoziţia 8.3.1. Funcţia $d_1 : \mathcal{D}[\mathcal{S}] \times \mathcal{D}[\mathcal{S}] \to [0, \infty]$ definită de relaţia

$$d_1(\mathcal{T}_j, \mathcal{T}_k) = H(\mathcal{T}_j/\mathcal{T}_k) + H(\mathcal{T}_k/\mathcal{T}_j) \tag{8.3.1}$$

este metrică în $\mathcal{D}[\mathcal{S}]$.

Demonstraţie. Avînd în vedere proprietăţile entropiei, primele două axiome ale distanţei sînt evidente. Pentru inegalitatea triunghiului avem

$$d_1(\mathcal{T}_j, \mathcal{T}_l) + d_1(\mathcal{T}_l, \mathcal{T}_k) = H(\mathcal{T}_j/\mathcal{T}_l) + H(\mathcal{T}_l/\mathcal{T}_j) + H(\mathcal{T}_l/\mathcal{T}_k) + H(\mathcal{T}_k/\mathcal{T}_l). \tag{8.3.2}$$

Scanned with OKEN Scanner

Cum însă din paragraful 3.3 avem

$$H(\mathcal{T}_j/\mathcal{T}_l) \geqslant H[\mathcal{T}_j/(\mathcal{T}_k, \mathcal{T}_l)] =$$

$$= H([(\mathcal{T}_j, \mathcal{T}_l)/\mathcal{T}_k] - (\mathcal{T}_l/\mathcal{T}_k) \qquad (8.3.3)$$

$$H(\mathcal{T}_k/\mathcal{T}_l) \geqslant H[\mathcal{T}_k/(\mathcal{T}_j, \mathcal{T}_l)] =$$

$$= H[(\mathcal{T}_k, \mathcal{T}_l)/\mathcal{T}_j] - H(\mathcal{T}_l/\mathcal{T}_j) \qquad (8.3.4)$$

rezultă că

$$d_1(\mathcal{T}_j, \mathcal{T}_l) + d_1(\mathcal{T}_l, \mathcal{T}_k) \geqslant H[(\mathcal{T}_j, \mathcal{T}_l/\mathcal{T}_k] + H[(\mathcal{T}_k, \mathcal{T}_l)/\mathcal{T}_j]. \qquad (8.3.5)$$

Dar pentru orice $\mathcal{T}_j \in \mathcal{D}[\mathcal{S}]$

$$H[(\mathcal{T}_j, \mathcal{T}_l)/\mathcal{T}_k] \geqslant H(\mathcal{T}_j/\mathcal{T}_k) \text{ și } H[(\mathcal{T}_k, \mathcal{T}_l)/\mathcal{T}_j] \geqslant H(\mathcal{T}_k/\mathcal{T}_j) \qquad (8.3.6)$$

și prin urmare

$$d_1(\mathcal{T}_j, \mathcal{T}_l) + d_1(\mathcal{T}_l, \mathcal{T}_k) \geqslant H(\mathcal{T}_j/\mathcal{T}_k) +$$

$$+ H(\mathcal{T}_k/\mathcal{T}_j) = d_1(\mathcal{T}_j, \mathcal{T}_k) \qquad (8.3.7)$$

și propoziția este demonstrată.

Metrica (8.3.1) a fost introdusă de C. E. Shannon în 1948. Să remarcăm că pentru metrica (8.3.1) axioma a) este satisfăcută într-un caz mult mai general și anume $d(\mathcal{T}_j, \mathcal{T}_k) = 0$, dacă și numai dacă, $\mathcal{T}_j$ r.c.d. $\mathcal{T}_k$, fiind vorba de fapt de repartițiile probabiliste ale lui $\mathcal{T}_j$ și $\mathcal{T}_k$. Evident că dacă $\mathcal{T}_j = \mathcal{T}_k$ atunci $\mathcal{T}_j$ r.c.d. $\mathcal{T}_k$ dar nu și reciproc.

Pusă sub o altă formă $d_1(\mathcal{T}_j, \mathcal{T}_k)$ are expresia

$$d_1(\mathcal{T}_j, \mathcal{T}_k) = 2H(\mathcal{T}_j, \mathcal{T}_k) - [H(\mathcal{T}_j) + H(\mathcal{T}_k)] \qquad (8.3.8)$$

sau încă

$$d_1(\mathcal{T}_j, \mathcal{T}_k) = H(\mathcal{T}_j, \mathcal{T}_k) - W[\mathcal{T}_j, \mathcal{T}_k]. \qquad (8.3.9)$$

Deoarece dacă d_1 este metrică atunci și λd_1 este metrică pentru orice constantă reală pozitivă λ, punînd $\lambda = \frac{1}{2}$,

din (8.3.8) deducem că

$$\frac{1}{2}\, d_1(\mathcal{T}_j, \mathcal{T}_k) = \frac{1}{2}\,[H(\mathcal{T}_j) + H(\mathcal{T}_k)] - W[\mathcal{T}_j, \mathcal{T}_k] \quad (8.3.10)$$

este de asemenea o metrică în $\mathcal{D}[\mathcal{S}]$. Cum $H(\mathcal{T}_j, \mathcal{T}_k)$ şi $\frac{1}{2}\,[H(\mathcal{T}_j) + H(\mathcal{T}_k)]$ sînt doi dintre majoranţii lui $W[\mathcal{T}_j, \mathcal{T}_k]$ din relaţia (7.6.23) pentru $m = 2$, observăm că luînd ceilalţi doi majoranţi, expresiile

$$\max\{H(\mathcal{T}_j),\ H(\mathcal{T}_k)\} - W[\mathcal{T}_j, \mathcal{T}_k] \text{ şi } \min\{H(\mathcal{T}_j),\ H(\mathcal{T}_k)\} - W[\mathcal{T}_j, \mathcal{T}_k] \quad (8.3.11)$$

nu mai sînt metrici în $\mathcal{D}(\mathcal{S})$. În schimb suma lor este chiar expresia (8.3.8) deci metrica $d_1(\mathcal{T}_j, \mathcal{T}_k)$.

Propoziţia 8.3.2. Funcţia $d_2 : \mathcal{D}[\mathcal{S}] \times \mathcal{D}[\mathcal{S}] \to [0,1]$ definită de relaţia

$$d_2(\mathcal{T}_j, \mathcal{T}_k) = \frac{H(\mathcal{T}_j/\mathcal{T}_k) + H(\mathcal{T}_k/\mathcal{T}_j)}{H(\mathcal{T}_j, \mathcal{T}_k)} \quad (8.3.12)$$

este metrică în $\mathcal{D}[\mathcal{S}]$.

Demonstraţie. Axiomele a) şi b) ale distanţei sînt evidente. Pentru a demonstra axioma triunghiului ţinem cont din nou de proprietăţile entropiei prezentate în paragraful 3.3 şi în plus de inegalitatea evidentă

$$\frac{a}{a+b} \geqslant \frac{c}{c+b} \quad (8.3.13)$$

pentru orice $a, b, c > 0$, $a \geqslant c$. Să scriem

$$d_2(\mathcal{T}_j, \mathcal{T}_k) = \left[\frac{H(\mathcal{T}_j/\mathcal{T}_l)}{H(\mathcal{T}_j, \mathcal{T}_l)} + \frac{H(\mathcal{T}_l/\mathcal{T}_k)}{H(\mathcal{T}_k, \mathcal{T}_l)}\right] + \left[\frac{H(\mathcal{T}_l/\mathcal{T}_j)}{H(\mathcal{T}_j, \mathcal{T}_l)} + \frac{H(\mathcal{T}_k/\mathcal{T}_l)}{H(\mathcal{T}_l, \mathcal{T}_k)}\right]. \quad (8.3.14)$$

Deoarece

$$H(\mathcal{T}_l) \leqslant H(\mathcal{T}_k) + H(\mathcal{T}_l/\mathcal{T}_k) \leqslant H(\mathcal{T}_k) + H(\mathcal{T}_l/\mathcal{T}_k) + H(\mathcal{T}_j/\mathcal{T}_l) \quad (8.3.15)$$

Scanned with OKEN Scanner

deducem că

$$\frac{H(\mathcal{T}_j/\mathcal{T}_l)}{H(\mathcal{T}_j, \mathcal{T}_l)} + \frac{H(\mathcal{T}_l/\mathcal{T}_k)}{H(\mathcal{T}_l, \mathcal{T}_k)} \geqslant$$

$$\geqslant \frac{H(\mathcal{T}_j/\mathcal{T}_l) + H(\mathcal{T}_l/\mathcal{T}_k)}{H(\mathcal{T}_k) + H(\mathcal{T}_j/\mathcal{T}_l) + H(\mathcal{T}_l/\mathcal{T}_k)}. \qquad (8.3.16)$$

Din (8.3.3) rezultă că $H(\mathcal{T}_j/\mathcal{T}_l) + H(\mathcal{T}_l/\mathcal{T}_k) \geqslant H(\mathcal{T}_j/\mathcal{T}_k)$ și ținînd cont de (8.3.13) deducem că

$$\frac{H(\mathcal{T}_j/\mathcal{T}_l)}{H(\mathcal{T}_j, \mathcal{T}_l)} + \frac{H(\mathcal{T}_l/\mathcal{T}_k)}{H(\mathcal{T}_l, \mathcal{T}_k)} \geqslant$$

$$\geqslant \frac{H(\mathcal{T}_j/\mathcal{T}_k)}{H(\mathcal{T}_k) + H(\mathcal{T}_j/\mathcal{T}_k)} = \frac{H(\mathcal{T}_j/\mathcal{T}_k)}{H(\mathcal{T}_j, \mathcal{T}_k)}. \qquad (8.3.17)$$

Inversînd j cu k în (8.2.17) rezultă

$$\frac{H(\mathcal{T}_k/\mathcal{T}_l)}{H(\mathcal{T}_k, \mathcal{T}_l)} + \frac{H(\mathcal{T}_l/\mathcal{T}_j)}{H(\mathcal{T}_l, \mathcal{T}_j)} \geqslant \frac{H(\mathcal{T}_k/\mathcal{T}_j)}{H(\mathcal{T}_k, \mathcal{T}_j)} \qquad (8.3.18)$$

și însumînd (8.3.17) și (8.3.18) termen cu termen rezultă că

$$d_2(\mathcal{T}_j, \mathcal{T}_l) + d_2(\mathcal{T}_l, \mathcal{T}_k) \geqslant d_2(\mathcal{T}_j, \mathcal{T}_k) \qquad (8.3.19)$$

adică tocmai inegalitatea triunghiului și propoziția este demonstrată.

Metrica (8.3.12) a fost introdusă de Rajski în 1961.

Remarcăm că atît $d_1(\mathcal{T}_j, \mathcal{T}_k)$ cît și $d_2(\mathcal{T}_j, \mathcal{T}_k)$ își ating valorile maxime în cazul în care subsistemele sînd independente. Se poate verifica de asemenea ușor

Propoziția 8.3.3. Funcțiile $D_k: \mathfrak{D}[\mathcal{S}] \times \mathfrak{D}[\mathcal{S}] \to (0, \infty)$ definite de relația

$$D_k(\mathcal{T}_j, \mathcal{T}_k) = \inf_{\substack{B_i \in \mathcal{T}_i \\ B_j \in \mathcal{T}_j}} \{d_k(B_i, B_j)\} \qquad k = 1,2 \qquad (8.3.20)$$

sînt de asemenea metrici în $\mathfrak{D}[\mathcal{S}]$.

În baza celor de mai sus remarcăm de asemenea că metricile introduse constituie ele însele măsuri ale gradului de dependență sau de independență dintre subsistemele $\mathcal{T}_j$ și $\mathcal{T}_k$, fiind de altfel complementare față de $H(\mathcal{T}_j, \mathcal{T}_k)$ și res-

Scanned with OKEN Scanner

pectiv unu, ale măsurilor globale $W[\mathcal{T}_j, \mathcal{T}_k]$ și $G_1[\mathcal{T}_j, \mathcal{T}_k]$. Interdependența dintre două subsisteme poate fi privită acum și prin distanța entropică dintre ele. Cu cît distanța este mai mare, cu atît gradul de interdependență este mai mic și reciproc.

Să presupunem acum că $\mathcal{S}$ este un sistem de comunicație. Atunci $d_1(\mathcal{T}_1, \mathcal{T}_3)$, în baza relației (8.3.1), reprezintă distanța dintre emisie și recepție, prin prizma cantității de informație care se pierde pe canal. Avînd spre exemplu un sistem de comunicație în dublu sens, deci în care $\mathcal{T}_1$ și $\mathcal{T}_3$ sînt pe rînd și surse de emisie și surse de recepție, cantitatea totală de informație care se pierde în ambele sensuri este $H(\mathcal{T}_1/\mathcal{T}_3) + H(\mathcal{T}_3/\mathcal{T}_1)$ adică tocmai $d_1(\mathcal{T}_1, \mathcal{T}_3)$. Apare ca natural faptul că distanța informațională dintre emisie și recepție este zero dacă și numai dacă pe canal nu se pierde nimic, ceea ce se întîmplă dacă și numai dacă emisia și recepția sînt reciproc complet dependente. Nepierzîndu-se nici o informație pe canal, putem, din acest punct de vedere, să considerăm că emisia și recepția sînt identice. Distanța dintre sursa de emisie și cea de recepție este maximă atunci cînd pe canal se pierde întreaga cantitate de informație care s-a emis, ceea ce se întîmplă atunci cînd perturbația este maximă și cînd practic emisia și recepția sînt ca și independente una de ala. Ca urmare putem aprecia că relațiile (8.3.1) și (8.3.12) reprezintă măsuri ale distanței sau depărtării între ceea ce informațional s-a emis și ceea ce s-a recepționat. Cu cît această depărtare este mai mică, cu atît transmisia este mai bună.

Pornind de la aceste considerente am putea introduce și conceptul de lungime entropică a unui canal de comunicație definit de expresia

$$L(\mathcal{T}_2) = H(\mathcal{T}_1/\mathcal{T}_3) + H(\mathcal{T}_3/\mathcal{T}_1) \qquad (8.3.21)$$

care, analitic, coincide cu distanța de la emisie la recepție $d_1(\mathcal{T}_1, \mathcal{T}_3)$. Evident că $L(\mathcal{T}_2) = 0 \Leftrightarrow d_1(\mathcal{T}_1, \mathcal{T}_3) = 0$ și $L(\mathcal{T}_2) = H(\mathcal{T}_1) + H(\mathcal{T}_2) \Leftrightarrow d_1(\mathcal{T}_1, \mathcal{T}_3) = H(\mathcal{T}_1) + H(\mathcal{T}_3)$. Ca urmare, vom spune că transmisia este fără perturbații sau că este perfectă, cînd lungimea entropică a canalului de comunicații este zero și că transmisia este total perturbată dacă lungimea entropică a canalului este maximă, avînd totdeauna $0 \leqslant L(\mathcal{T}_2) \leqslant H(\mathcal{T}_1) + H(\mathcal{T})_3$.

În mod analog putem introduce și lungimea entropică normată a canalului de comunicație, avînd expresia

$$l(\mathcal{T}_2) = d_2(\mathcal{T}_1, \mathcal{T}_3) \qquad (8.3.22)$$

Scanned with OKEN Scanner

avînd transmisie perfectă dacă $l(\mathcal{T}_2) = 0$ și transmisie total perturbată dacă $l(\mathcal{T}_2) = 1$.

Ținînd cont de lungimea entropică normată a canalului de comunicație putem introduce și conceptul de fidelitate entropică a unui canal de comunicație a cărui expresie este

$$\Phi(\mathcal{T}_2) = 1 - l(\mathcal{T}_1, \mathcal{T}_3) \qquad (8.3.23)$$

ceea ce ne-ar permite să spunem că avem o fidelitate maximă, sau o transmisie perfectă dacă $\Phi(\mathcal{T}_2) = 1$ și o transmisie total perturbată cînd fidelitatea este zero. Convențional, am putea spune că transmisia este slabă ca fidelitate dacă $\Phi(\mathcal{T}_2) < 0{,}25$, că transmisia are o fidelitate bună sau puternică dacă $0{,}25 \leqslant \Phi(\mathcal{T}_2) \leqslant 0{,}75$ și că avem o înaltă fidelitate entropică sau informațională dacă $\Phi(\mathcal{T}_2) > 0{,}75$.

Presupunînd acum că $\mathcal{S}$ este un sistem de strategii aleatoare, expresia $d_1(\mathcal{C}_j, \mathcal{C}_k)$ poate fi interpretată ca distanță informațională între două coaliții. Aceasta este zero în cazul în care cele două coaliții coincid informațional sau, cu alte cuvinte, în cazul în care din punct de vedere probabilistic, cele două coaliții adoptă aceeași strategie. Distanța dintre cele două coaliții devine maximă în cazul în care din punct de vedere probabilistic acestea adoptă strategii de joc independente, fiecare din ele neținînd cont de jocul celeilalte.

În particular, $d_1(\mathcal{C}_j, \mathcal{C}_k)$ poate reprezenta, prin prisma strategiilor de joc adoptate, distanța dintre doi jucători sau dintre un jucător și o coaliție. Indiferent de situație, distanța informațională este cu atît mai mică cu cît interdependența dintre jucători este mai mare, ceea ce apare de asemenea ca natural și subliniază calitatea măsurilor entropice în raport cu celelalte măsuri ale dependenței de a reliefa această informație remarcabilă.

9. Organizare

9.1. Măsuri entropice ale organizării

În sensul cel mai larg al noțiunii de organizare, apare ca intuitiv faptul că într-un sistem aleator $\mathcal{S}$ ale cărui elemente sînt legate între ele există o anumită organizare, în virtutea căreia se poate vorbi de o anumită structură organizatorică a sistemului. Ca urmare, putem studia intensitatea cone-

Scanned with OKEN Scanner

xiunilor dintre diferite părți ale lui $\mathcal{S}$ și deducem ceea ce se poate întîmpla cînd sistemului i se adaugă noi elemene sau i se elimină unele elemente constitutive. Evident că pe măsură ce blocurile sistemului tind spre a nu mai depinde unele de altele, nedeterminarea crește, devenind maximă odată cu absența oricărei legături dintre blocuri, deci cînd acestea sînt independente. Dacă fiecare element al sistemului evoluează independent de restul sistemului, rezultă că la nivel de ansamblu nu poate fi vorba de nici o legătură între părțile lui S, ceea ce implică absența oricărei organizări. Dacă însă în evoluția lor blocurile tind spre a fi din ce în ce mai legate unele de altele, ajungînd la situația extremă de a fi complet dependente unele de altele atunci nedeterminarea descrește, devenind zero în cazul extrem al dependenței complete, cînd putem ști totul despre sistem, ceea ce implică o organizare perfectă. Cum între organizare și nedeterminare există o legătură naturală și cum pentru nedeterminare există deja o măsură, pornind de la aceasta se poate soluționa și problema măsurării organizării unui sistem aleator $\mathcal{S}$. Conceptul de organizare a unui sistem aleator $\mathcal{S}$, așa cum arată S. Guiașu (1973), reflectă faptul că „cu cît un sistem este mai organizat, cu atît mai bine este definită comportarea ansamblului în ciuda comportării nedeterminate a indivizilor care alcătuiesc sistemul". În baza acestei sublinieri și a legăturii dintre conceptele de nedeterminare și organizare în cazul unui sistem aleator $\mathcal{S}$, S. Watanabe (1969) și S. Guiașu (1973) au propus drept măsură a organizării unui sistem tocmai diferența dintre nedeterminarea părților sistemului și nedeterminarea ansamblului. Fie așadar sistemul $\mathcal{S}$ cu desfacerea $\mathcal{D}[\mathcal{S}]$ dată de relația (8.1.1,).

Definiția 9.1.1. Se numește măsură a organizării sistemului $\mathcal{S}$ cu structura $[\mathcal{D}, Q_{\mathcal{D}}[\mathcal{S}]]$ sau pe scurt, organizare W—G a sistemului $\mathcal{S}$, expresia

$$W[\mathcal{S}; \mathcal{D}[\mathcal{S}]] = \sum_{j=1}^{m} H(\mathcal{T}_j) - H(\mathcal{S}). \qquad (9.1.1)$$

Să remarcăm că analitic, relațiile (8.2.1) și (9.1.1) coincid, interpretarea lor fiind alta, ambele fiind de altfel cazuri particulare de utilizare și interpretare concretă și adecvată a relației (7.6.19). Ca urmare, proprietățile măsurii (7.6.19) rămîn valabile în întregime și pentru organizarea W—G (9.1.1), cu semnificațiile corespunzătoare.

Scanned with OKEN Scanner

Definiția 9.1.2. Vom spune că sistemul $\mathcal{S}$ are o structură organizatorică de independență globală dacă elementele lui $\mathfrak{D}[\mathcal{S}]$ sînt global independente și că acesta are o structură organizatorică de dependență totală dacă elementele lui $\mathfrak{D}[\mathcal{S}]$ sînt global reciproc complet dependente.

Ținînd cont de considerațiile din paragraful 8.1, dependența dintre elementele lui $\mathfrak{D}[\mathcal{S}]$ apare ca dependență dintre vectori aleatori și ca urmare a propozițiilor (7.6.9) — (7.6.11) putem enunța și deduce

Propoziția 9.1.1. Pentru măsura organizării W — G $W[\mathcal{S}; \mathfrak{D}[\mathcal{S}]]$ sînt adevărate întotdeauna următoarele:

1) este definită pentru orice sistem aleator $\mathcal{S}$ și orice structură $[\mathfrak{D}, Q_{\mathfrak{D}}[\mathcal{S}]]$ a acestuia;

2) este simetrică în raport cu elementele lui $\mathfrak{D}[\mathcal{S}]$;

3) $0 \leqslant W[\mathcal{S}; \mathfrak{D}[\mathcal{S}]] < \infty$;

4) $W[\mathcal{S}; \mathfrak{D}[\mathcal{S}]] = 0$ dacă și numai dacă $\mathcal{S}$ are o structură de independență globală;

5) $W[\mathcal{S}; \mathfrak{D}[\mathcal{S}]] = \sum_{k=2}^{m} [H(\mathcal{T}_{j_k}) - H[\mathcal{T}_{j_k}/(\mathcal{T}_{j_1}, \ldots, \mathcal{T}_{j_{k-1}})]]$ pentru orice permutare a elementelor lui $\mathfrak{D}[\mathcal{S}]$;

6) $W[\mathcal{S}, \mathfrak{D}[\mathcal{S}]] \geqslant \sum_{k=2}^{m} [H(\mathcal{T}_{j_k}) - H(\mathcal{T}_{j_k}/T_{j_{k-1}})]$ cu egalitate dacă și numai dacă există o permutare a subsistemelor lui $\mathfrak{D}[\mathcal{S}]$ astfel încît $\{\mathcal{T}_{j_1}, \mathcal{T}_{j_2}, \ldots, \mathcal{T}_{j_{k-2}}\}$ c.d. $\mathcal{T}_{j_{k-1}}$, $3 \leqslant k \leqslant m$;

7) $W[\mathcal{S}; \mathfrak{D}[\mathcal{S}]] \leqslant \sum_{k=1}^{m-1} H(\mathcal{T}_{j_k})$ cu egalitate dacă și numai dacă există o permutare a elementelor lui $\mathfrak{D}[\mathcal{S}]$ astfel încît $\mathcal{T}_{j_k}$ să fie complet dependent de cel puțin numai unul din subsistemele $\mathcal{T}_{j_1}, \mathcal{T}_{j_2}, \ldots, \mathcal{T}_{j_{k-1}}$, $2 \leqslant k \leqslant m$;

8) $W[\mathcal{S}; \mathfrak{D}[\mathcal{S}]] = (m-1)H(\mathcal{T}_s) = (m-1)H(\mathcal{S})$ dacă și numai dacă pentru orice permutare a subsistemelor lui $\mathfrak{D}[\mathcal{S}]$ avem $\{\mathcal{T}_{j_1}, \mathcal{T}_{j_2}, \ldots, T_{j_{k-2}}\}$ c.d. $T_{j_{k-1}}$, $3 \leqslant k \leqslant m$, $1 \leqslant s \leqslant m$;

9) $W[\mathcal{S}; B_1, B_2, \ldots, B_n] \geqslant W[\mathcal{S}, \mathfrak{D}[\mathcal{S}]]$ pentru orice desfacere neatomică, cu egalitate dacă și numai dacă blocurile lui $\mathcal{S}$ sînt global independente;

Scanned with OKEN Scanner

10) $W[\mathcal{S}; B_1, B_2, \ldots, B_n] \geqslant \sum_{j=1}^{m} W[\mathcal{T}_j; \mathcal{D}[\mathcal{T}_j]]$ pentru orice $\mathcal{D}[\mathcal{S}]$ și orice desfaceri ale subsistemelor lui $\mathcal{D}[\mathcal{S}]$ cu egalitate în cazurile extreme ale dependenței globale sau ale independenței globale;

11) este invariantă în raport cu orice transformare biunivocă a lui $\mathcal{S}$.

În mod analog putem formula în termenii organizării și relațiile (7.6.23) — (7.6.36). Ca urmare a relației (7.6.28) putem defini coeficienții de organizare entropică ai sistemului $\mathcal{S}$ cu structura $[\mathcal{D}\, Q_{\mathcal{D}}[\mathcal{S}]]$ ca fiind

$$G_i[\mathcal{S}; \mathcal{D}[\mathcal{S}]] = \frac{W[\mathcal{S}; \mathcal{D}[\mathcal{S}]]}{M_i[\mathcal{D}[\mathcal{S}]]}, \quad 1 \leqslant i \leqslant 4 \qquad (9.1.2)$$

pentru care propozițiile (7.6.14) — (7.6.16) rămîn valabile cu terminologia de organizare a unui sistem.

Să presupunem acum că sistemul $\mathcal{S}$ este constituit parte cu parte, prin juxtapunerea subsistemelor ce determină $\mathcal{D}[\mathcal{S}]$, unul cîte unul, ca în (7.6.32). Fără a diminua generalitatea problemei, să admitem că juxtapunerea se face în ordinea naturală $\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_m$. Dacă $\mathcal{T}_{k+1}$ c.d. $\mathcal{T}_k$, $1 \leqslant k \leqslant m-1$, atunci, în baza proprietății 7) deducem că $W[\mathcal{S}; \mathcal{T}_1, \ldots \ldots \mathcal{T}_m] = \sum_{k=2}^{m} H(\mathcal{T}_k)$. Dacă $\{\mathcal{T}_1, \ldots, \mathcal{T}_{k-2}\}$ c.d. $\mathcal{T}_{k-1}$, $3 \leqslant k \leqslant m$, atunci $W[\mathcal{S}; \mathcal{T}_1, \ldots, \mathcal{T}_m] = \sum_{k=2}^{m} [H(\mathcal{T}_k) - H(\mathcal{T}_k/\mathcal{T}_{k-1})]$ în baza proprietății 6), iar dacă în plus $\mathcal{T}_k$ este independent de $\mathcal{T}_{k-1}$, $2 \leqslant k \leqslant m$, atunci $W[\mathcal{S}; \mathcal{T}_1, \ldots, \mathcal{T}_m] = 0$. În cazul mai general al subsistemelor lui $\mathcal{S}$, relația (8.6.27) capătă o semnificație remarcabilă, avînd, la constituirea lui $\mathcal{S}$

$$W[\mathcal{D}[\mathcal{S}]] = W[\mathcal{D}[\mathcal{S}] \setminus \{\mathcal{T}\}] + W[\mathcal{D}[\mathcal{S}] \setminus \{\mathcal{T}\}, \mathcal{T}] \qquad (9.1.3)$$

ceea ce demonstrează.

Propoziția 9.1.2. Dacă unui sistem i se mai adaugă un subsistem, atunci:

1) gradul de organizare a noului sistem este dat de relația (9.1.3);

2) noul sistem are un grad de organizare superior vechiului sistem, cu egalitate dacă și numai dacă subsistemul adăugat este independent de sistemul la care s-a adăugat.

Ca urmare, dacă $\mathcal{D}_1[\mathcal{S}] \subseteq \mathcal{D}_2[\mathcal{S}]$, sînt două etape succesive de constituire a lui $\mathcal{S}$, $\mathcal{D}_2[\mathcal{S}] = \mathcal{D}_1[\mathcal{S}] \cup \{\mathcal{T}\}$ atunci $W[\mathcal{D}_1[\mathcal{S}]] \leqslant W[\mathcal{D}_2[\mathcal{S}]]$ ceea ce demonstrează

Propoziția 9.1.3. Măsura organizării $W - G$, $W[\mathcal{S}; \mathcal{D}[\mathcal{S}]]$ este monoton crescătoare pe mulțimea $\{\mathcal{D}_k[\mathcal{S}]/\mathcal{D}_k[\mathcal{S}] \subseteq \subseteq \mathcal{D}_{k+1}[\mathcal{S}],\ 1 \leqslant k \leqslant m,\ \mathcal{D}_m[\mathcal{S}] = \mathcal{S}\}$ adică $W[\mathcal{D}_k[\mathcal{S}]] \leqslant \leqslant W[\mathcal{D}_{k+1}[\mathcal{S}]],\quad 1 \leqslant k \leqslant m.$

Dacă relația (9.1.3) reprezintă „agregarea" organizării corespunzătoare constituirii sistemului $\mathcal{S}$, atunci, scriind-o sub forma

$$W[\mathcal{D}[\mathcal{S}] \setminus \{\mathcal{T}\}] = W[\mathcal{D}[\mathcal{S}]] - W[\mathcal{D}[\mathcal{S}] \setminus \{\mathcal{T}\}, \mathcal{T}]. \qquad (9.1.4)$$

deducem un rezultat analog celui din propoziția 9.1.2, dar pentru cazul în care sistemul $\mathcal{S}$ este desfăcut parte cu parte, după elementele lui $\mathcal{D}[\mathcal{S}]$.

În sensul conceptului de organizare $W - G$ relațiile (7.6.34) pot defini măsurile organizării totale $W_{tot}[\mathcal{S}]$, $W_{tot}[\mathcal{D}[\mathcal{S}])$ și $W_{tot}[\mathcal{T}]$, respectiv pentru sistemul $\mathcal{S}$ cu desfacerea atomică, cu desfacerea $\mathcal{D}[\mathcal{S}]$ neatomică și pentru subsistemul $\mathcal{T}$ cu desfacere atomică. În acest context, propoziția 7.6.17 capătă o semnificație aparte conducînd la

Propoziția 9.1.4. Suma tuturor organizărilor de forma (9.1.1) corespunzătoare desfacerilor sau constituirilor consecutive ale unui sistem $\mathcal{S}$ este independentă de modul și numărul de etape în care acestea se fac și este egală cu $W_{tot}[\mathcal{S}]$.

În baza propoziției 9.1.4, dacă privim procedeul de constituire sau desfacere a lui $\mathcal{S}$ ca o schemă de tip arborescență, atunci fiecărui nod al acesteia îi corespunde o măsură a organizării de tip (9.1.1) sau pe scurt o organizare, a căror însumare după toate nodurile este egală cu $W_{tot}[\mathcal{S}] = \sum_{i=1}^{n} H(B_i) - H(\mathcal{S})$, indiferent de modul în care s-au ales elementele din fiecare nod și indiferent de numărul de noduri. Rezultatul, remarcabil de altfel, poate fi prezentat în mod analog și prin intermediul relației (7.6.26). Ținînd cont de cele de mai sus cît și de relația (7.6.35) putem enunța

Scanned with OKEN Scanner

Propoziția 9.1.5. Pentru orice sistem

$$1)\ W_{tot}[\mathcal{S}] = \max_{\mathfrak{D}} \{W_{tot}[\mathfrak{D}[\mathcal{S}]]\} \tag{9.1.5}$$

$$2)\ W_{tot}[\mathcal{S}] \geqslant \sum_{\mathcal{T} \in \mathfrak{D}[\mathcal{S}]} W_{tot}[\mathcal{S}] \tag{9.1.6}$$

$$3)\ W_{tot}[\mathcal{S}] \geqslant W_{tot}[\mathcal{T}], \quad (\forall)\ \mathcal{T} \subseteq \mathcal{S} \tag{9.1.7}$$

Din propoziția 9.1.5 deducem ca remarcabil faptul că dintre toate desfacerile sistemului $\mathcal{S}$, organizarea maximă o are desfacerea atomică. Ca o consecință imediată a propoziției 9.1.5 putem enunța

Propoziția 9.1.6. Pentru orice sistem $\mathcal{S}$ cu desfacerea $\mathfrak{D}[\mathcal{S}]$ avem :

1) dacă $W_{tot}[\mathcal{S}] = 0$ atunci $W_{tot}[\mathcal{T}] = 0$ pentru orice $\mathcal{T} \in \mathfrak{D}[\mathcal{S}]$

2) dacă există $\mathcal{T} \in \mathfrak{D}[\mathcal{S}]$ astfel că $W_{tot}[\mathcal{T}] > 0$, atunci $W_{tot}[\mathcal{S}] > 0$, reciprocele celor două afirmații fiind în general false.

Semnificația propoziției 9.1.6 rezultă din interpretarea ei Astfel, din afirmația 1) deducem că dacă sistemul $\mathcal{S}$ este lipsit de organizare în totalitate, atunci orice parte a acestuia este fără organizare, iar din afirmația 2) deducem că dacă cel puțin o parte a sistemului are o anumită organizare, atunci sistemul însuși are un anumit grad de organizare.

Să presupunem acum că $\mathcal{S}$ este un sistem de comunicații ca în paragrafele precedente. Remarcăm că vorbind despre organizarea cuplului emisie-recepție $(\mathcal{T}_1, \mathcal{T}_3)$, atunci măsura organizării acestuia (9.1.1) este tocmai viteza de transmitere a informației prin canal (8.2.15) pentru o repartiție anume a emisiei, sau capacitatea canalului (8.2.16), pentru repartiția care maximizează viteza de transmisie. În aceste condiții, pentru sistemele de comunicație cu emisie de tip Bernoulli și canal de comunicație staționar fără memorie și simetric, relația (8.2.17) dă chiar mărimea gradului de organizare al cuplului $(\mathcal{T}_1, \mathcal{T}_3)$.

Pentru același cuplu $(\mathcal{T}_1, \mathcal{T}_3)$ deducem fără dificultate

Propoziția 9.1.7. Dacă evoluția cuplului $(\mathcal{T}_1, \mathcal{T}_3)$ al unui sistem de comunicație cu emisie de tip Bernoulli și canal staționar fără memorie este descrisă de legea de repartiție

(7.2.1) atunci pentru orice $x \in A$, $y \in A$

$$1)\ p(x) = \sqrt{\frac{a_1 a_2 - h^2}{2\pi a_2}} \exp\left\{-\frac{a_1 a_2 - h^2}{a_2} x^2\right\},\ (\forall)\ x \in A \tag{9.1.8}$$

$$2)\ p(y) = \sqrt{\frac{a_1 a_2 - h^2}{2\pi a_1}} \exp\left\{-\frac{a_1 a_2 - h^2}{a_1} y^2\right\},\ (\forall)\ y \in A \tag{9.1.9}$$

$$3)\ p(x/y) =$$

$$= \sqrt{\frac{a_2}{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{h^2 - a_1 a_2}{a_2} x^2 + 2hxy + a^2 y^2\right)\right\},\ (\forall) x \in A,\ y \in B \tag{9.1.10}$$

$$4)\ p(y/x) =$$

$$= \sqrt{\frac{a_1}{2\pi}} \exp\left\{-\frac{1}{2}\left(a_1 x^2 + 2hxy + \frac{h^2 - a_1 a_2}{a_1} y^2\right)\right\},\ (\forall) x \in A,\ y \in B \tag{9.1.11}$$

$$5)\ W[\mathcal{T}_1, \mathcal{T}_2] = \frac{1}{2} \log \frac{a_1 a_2}{a_1 a_2 - h^2}. \tag{9.1.12}$$

În cazul particular în care a_1, a_2 și h sînt date de relațiile (7.6.46) repartiția (7.2.1) este repartiția normală bidimensională de medii zero, dispersii σ_1 și σ_2 și coeficient de corelație ρ. Ca urmare a propoziției 9.17 deducem

Propoziția 9.1.8. Dacă $\mathcal{S}$ este un sistem de comunicație cu emisie de tip Bernoulli și canal staționar fără memorie pentru care evoluția cuplului emisie-recepție este descrisă de repartiția normală bidimensională $N(0, 0, \sigma_1, \sigma_2, \rho)$ atunci

1) cantitatea medie de informație ce se pierde pe canal este

$$H(\mathcal{T}_i/\mathcal{T}_j) = \log \sigma_i \sqrt{2\pi e (1 - \rho^2)}$$

$$i, j = 1, 2;\ i \neq j \tag{9.1.13}$$

2) gradul de organizare al cuplului este

$$W[\mathcal{T}_1, \mathcal{T}_2] = -\frac{1}{2} \log (1 - \rho^2). \tag{9.1.14}$$

Scanned with OKEN Scanner

Să remarcăm, în baza relației (9.1.13) că în general, într-un sistem de comunicație cu conexiune inversă, cantitatea de informație care se pierde pe canal nu este aceeași în ambele sensuri. În cazul sistemului din propoziție, dacă $\sigma_1 = \sigma_2$ atunci cantitate de informație care se pierde pe canal în ambele sensuri este aceeași.

Să considerăm acum sistemele de comunicații $\mathcal{S}_1, \mathcal{S}_2, \ldots$ $\ldots, \mathcal{S}_p$ în sensul paragrafului 8.1.1, pe care le vom numi sisteme simple și cu care, prin juxtapunere constituim un sistem compus $\mathcal{U} = \{\mathcal{S}_i\}_{i=1}^{p}$. Un sistem compus de transmitere a informației poate fi privit, în sens general, ca mulțimea tuturor surselor de emisie, a tuturor canalelor de comunicație și a tuturor surselor de recepție, aflate într-o anumită conexiune.

Un caz particular de sisteme de comunicație compuse îl reprezintă sistemele de comunicație cuplate în serie, cînd ieșirea unuia devine intrare pentru următorul, din care un subcaz îl constituie sistemele de comunicație pentru care ieșirea finală coincide cu intrarea inițială.

De asemenea un caz particular de sisteme de comunicație compus îl constituie sistemele cuplate în paralel avînd aceeași intrare inițială și aceeași ieșire finală, între acestea putînd avea alte intrări și ieșiri intermediare, a cărui schemă generală poate fi reprezentată printr-un graf orientat sau neorientat după cum este vorba despre un sistem fără sau cu conexiune inversă. În oricare din cazuri, presupunînd că $\mathcal{U} = \{\mathcal{S}_i\}_{i=1}^{p}$ este staționar și că repartițiile probalistice ce descriu evoluția sistemului compus sînt cunoscute sau determinabile, putem pune problema măsurii gradului de organizare $W[\mathcal{U}; \{\mathcal{S}_i\}_{i=1}^{p}]$ al sistemului compus $\mathcal{U}$ precum și a unor părți ale acestuia, considerațiile de mai sus rămînînd valabile în întregime.

Să presupunem acum că $\mathcal{S}$ reprezintă mulțimea partenerilor unui joc cu strategii aleatoare iar $\mathcal{D}[\mathcal{S}]$ o coalizare a acestora. În acest context relația (9.1.1), devenind

$$W[\mathcal{S}; \mathcal{D}[\mathcal{S}]] = \sum_{j=1}^{m} H(\mathcal{C}_j) - H(\mathcal{S}) \qquad (9.1.15)$$

reprezintă gradul de organizare $W - G$ al jocului cu n parteneri dispuși în $m < n$ coaliții, relație care, pentru $m = n$ devine

$$W_{tot}[\mathcal{S}] = \sum_{j=1}^{n} H(J_i) - H(\mathcal{S}) \qquad (9.1.16)$$

Scanned with OKEN Scanner

și reprezintă organizarea totală a jocului cu cei n parteneri necoalizați. În baza relației (9.1.3), formulată pentru un joc aleator, deducem că gradul de organizare al unui joc este crescător. Ca urmare, dacă la cele m coaliții $\{\mathcal{C}_j\}_{j=1}^{m}$ se mai adaugă una $\mathcal{C}_{m+1}$ formată din alți parteneri decît cei n considerați deja, gradul de organizare al celor $m+1$ coaliții este superior celui corespunzător celor m coaliții inițiale, cu excepția cazului în care noua coaliție este independentă de ansamblul celorlalte m. De asemenea, în baza relației (9.1.4) deducem că gradul de organizare al jocului scade prin retragerea din joc a unei coaliții, cu excepția cazului în care aceasta este independentă de coalițiile rămase. Propoziția 9.1.1 rămîne valabilă și capătă o semnificație interesantă în terminologia adecvată jocurilor aleatoare. Dacă $\mathfrak{D}_1[\mathcal{S}]$ este o coalizare a celor n jucători de forma (8.1.1), $\mathfrak{D}_2[\mathcal{S}]$ o coalizare de același tip dar aplicată fiecărei coaliții $\mathcal{C}_j \in \mathfrak{D}_1[\mathcal{S}]$ deci $\mathfrak{D}_2[\mathcal{S}] = \{\mathfrak{D}_1[\mathcal{C}_j]\}_{j=1}^{m} = \{\{\mathcal{C}_{jk}\}_{k=1}^{m_j}\}_{j=1}^{m}$ și așa mai departe pînă cînd fiecare coaliție este formată dintr-un partener, procedeul terminîndu-se în s etape, atunci propoziția 9.1.4 poate fi formulată astfel

Propoziția 9.1.9. Organizarea totală a jocului $W_{tot}[\mathcal{S}]$ este independentă de modul în care se coalizează cei n parteneri și

$$W_{tot}[\mathcal{S}] = \sum_{i=1}^{s} W[\mathcal{S}; \mathfrak{D}_i[\mathcal{S}]]. \tag{9.1.17}$$

Importanța acestei propoziții rezultă mai ales din consecințele ei, care pentru un sistem oarecare sînt de altfel formulate în propozițiile 9.1.5 și 9.1.6. În terminologia teoriei jocurilor acestea devin

Consecința 9.1.9.1. Dacă $\mathcal{C}$ este o coaliție netrivială a lui $\mathcal{S}$ atunci

$$W_{tot}[\mathcal{S}] = \max_{\mathcal{C} \subseteq \mathcal{S}} \{W_{tot}[\mathcal{C}]\} \tag{9.1.18}$$

Relația (9.1.18) spune că organizarea totală maximă corespunde jocului în care partenerii sînt necoalizați, ceea ce ilustrează un fapt de altfel natural și anume acela că interesul maxim al unui joc în măsurarea forțelor partenerilor de întrecere este atins atunci cînd se joacă fără coaliții. Relația (9.1.18) este întărită de

Scanned with OKEN Scanner

Consecința 9.1.9.2. Dacă coalizarea $\mathcal{D}[\mathcal{S}]$ este netrivială atunci

$$W_{tot}[\mathcal{S}] \geqslant \sum_{j=1}^{m} W_{tot}[\mathcal{C}_j] \qquad (9.1.19)$$

cu egalitate numai în cazurile extreme ale dependenței dintre coaliții.

Consecința 9.1.9.3. Pentru orice coalizare netrivială $\mathcal{D}[\mathcal{S}]$ a partenerilor

1) dacă $W_{tot}[\mathcal{S}] = 0$ atunci $W_{tot}[\mathcal{C}] = 0$ pentru orice $\mathcal{C} \subset \mathcal{S}$, adică, dacă pe ansamblu jocul este fără organizare, atunci orice coaliție este neorganizată,

2) dacă există $\mathcal{C} \subset \mathcal{S}$ cu $W_{tot}[\mathcal{C}] > 0$ atunci $W_{tot}[\mathcal{S}] > 0$, adică dacă cel puțin o coaliție are o anumită organizare, atunci la nivelul jocului există o anumită organizare, reciprocele celor două afirmații fiind în general false.

Pornind de la coeficienții de organizare (9.1.2) să notăm cu $G_{tot}[\mathcal{D}[\mathcal{S}]]$ coeficientul de organizare totală ca fiind unul dintre cei patru coeficienți definiți de relația (9.1.2) în cazul organizării totale $W_{tot}[\mathcal{S}]$.

Vom spune că jocul are o coalizare global independentă sau global fără organizare dacă $G_{tot}[\mathcal{D}[\mathcal{S}]] = 0$ și că acesta are o coalizare perfectă dacă $G_{tot}[\mathcal{D}[\mathcal{S}]] = 1$. De regulă, pentru majoritatea jocurilor $0 < G_{tot}[\mathcal{D}[\mathcal{S}]] < 1$. Am putea conveni să spunem că jocul are o coalizare slabă dacă $G_{tot}[\mathcal{D}[\mathcal{S}]] < 0{,}25$, o coalizare bună dacă $0{,}25 \leqslant G_{tot}[\mathcal{D}[\mathcal{S}]] \leqslant 0{,}75$ și o coalizare foarte puternică dacă $G_{tot}[\mathcal{D}[\mathcal{S}]] > 0{,}75$, cifrele luate fiind relative.

Măsuri entropice ale organizării pot fi definite și pe baza entropiei ponderate, și ca urmare, relația (8.2.12) poate defini măsura entropică ponderată a organizării sistemului $\mathcal{S}$ cu desfacerea $\mathcal{D}[\mathcal{S}]$, sau, pe scurt, organizarea ponderată $W-G$ a sistemului $\mathcal{S}$ cu desfacerea $\mathcal{D}[\mathcal{S}]$, măsură ale cărei proprietăți generale sînt date în paragraful 7.6.3 și care prezintă un interes particular prin faptul că organizarea este reflectată nu numai de cantitatea de informație corespunzătoare structurii sistemului ci și de utilitatea acestei informații.

Scanned with OKEN Scanner

9.2. Clasificare entropică

Fie A o mulțime oarecare, $\mathfrak{L}(A)$ familia tuturor submulțimilor lui A și $\mu : \mathfrak{L}(A) \to [0, \infty)$ o funcție de mulțime cu proprietatea că $\mu(\emptyset) = 0$.

Să considerăm orice două submulțimi disjuncte $A_1, A_2 \in \mathfrak{L}(A)$. Vom spune că funcția de mulțime

1) μ este aditivă dacă $\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2)$;

2) μ_- este subaditivă dacă $\mu_-(A_1 \cup A_2) \leqslant \mu_-(A_1) + \mu_-(A_2)$;

3) μ_+ este supraaditivă dacă $\mu_+(A_1 \cup A_2) \geqslant \mu_+(A_1) + \mu_+(A_2)$.

Vom spune că funcția de mulțime supraaditivă μ_+ și funcția de mulțime subaditivă μ_- sînt conjugate în raport cu funcția de mulțime aditivă μ dacă

$$\mu = \mu_+ + \mu_-. \tag{9.2.1}$$

Propoziția 9.2.1. Fie $\mathfrak{X}$ mulțimea tuturor variabilelor aleatoare discrete și $\mathfrak{L}(\mathfrak{X})$ familia tuturor părților acesteia. Atunci

1) funcția de mulțime

$$\mu(\mathcal{Y}) = \sum_{Y \in \mathcal{Y}} H(Y), \qquad \mathcal{Y} \subset \mathfrak{L}(\mathfrak{X}) \tag{9.2.2}$$

este aditivă;

2) funcția de mulțime

$$\mu_-(\mathcal{Y}) = H(\mathcal{Y}), \quad \mathcal{Y} \subset \mathfrak{L}(\mathfrak{X}) \tag{9.2.3}$$

este subaditivă;

3) funcția de mulțime

$$\mu_+(\mathcal{Y}) = W_{tot}[\mathcal{Y}], \quad \mathcal{Y} \subset \mathfrak{L}(\mathfrak{X}) \tag{9.2.4}$$

este supraaditivă;

4) funcțiile μ_+ și μ_- sînt conjugate în raport cu funcția μ.

Scanned with OKEN Scanner

Demonstrație. Dacă ținem cont de proprietățile entropiei din paragraful 3.3 și de cele ale corelației entropice globale din paragraful 7.6.1, afirmațiile sînt imediate.

Să presupunem acum că mulțimea A este un sistem aleator $\mathcal{S}$, iar $\mathfrak{A}(A)$ este o desfacere $\mathfrak{D}[\mathcal{S}]$ a acestuia. În linii mari, obținerea unei desfaceri $\mathfrak{D}[\mathcal{S}]$ reprezintă o grupare sau o etapă a unui procedeu de grupare a elementelor lui $\mathcal{S}$ după un anumit criteriu, sau pe scurt o clasificare a elementelor lui $\mathcal{S}$.

Ținînd cont de considerațiile din paragrafele 8.2 și 9.1 deducem imediat că prin descompunerea lui $\mathcal{S}$ interdependența sau coeziunea ce caracterizează sistemul ca un tot unitar, sau încă organizarea acestuia, scad, ceea ce implică o pierdere de coeziune sau de organizare în general. Prin urmare, prin desfacerea sistemului $\mathcal{S}$ în subsisteme „se plătește" o anumită „taxă", aceea a pierderii de coeziune, sau altfel spus clasificarea „costă".

Definiția 9.2.1. Fie $\mathfrak{D}[\mathcal{S}]$ o desfacere a lui $\mathcal{S}$. Se numește cost al desfacerii $\mathfrak{D}[\mathcal{S}]$ a lui $\mathcal{S}$ în sensul lui Watanabe, în raport cu o funcție de mulțime supraaditivă μ_+, expresia

$$C[\mathcal{S};\ \mathfrak{D}[\mathcal{S}]] = \mu_+(\mathcal{S}) - \sum_{\mathcal{T} \in \mathfrak{D}[\mathcal{S}]} \mu_+(\mathcal{T}) \qquad (9.2.5)$$

sau în raport cu funcția de mulțime subaditivă μ_- conjugată cu μ_+ în raport cu o funcție de mulțime aditivă μ, expresia

$$C[\mathcal{S};\ \mathfrak{D}[\mathcal{S}]] = \sum_{\mathcal{T} \in \mathfrak{D}[\mathcal{S}]} \mu_-(\mathcal{T}) - \mu_-(\mathcal{S}). \qquad (9.2.6)$$

Ca și în cazul interdependenței sau organizării sistemului $\mathcal{S}$ putem defini costul total al desfacerii lui $\mathcal{S}$ ca fiind

$$C_{tot}[\mathcal{S}] = \sum_{\substack{\text{(după toate} \\ \text{nodurile)}}} C\,(\text{corespunzător unui nod}) \qquad (9.2.7)$$

Să observăm că expresiile (9.2.5) și (9.2.6) coincid. Într-adevăr, scăzîndu-le, obținem

$$\mu_+(\mathcal{S}) + \mu_-(\mathcal{S}) - \left[\sum_{\mathcal{T} \in \mathfrak{D}[\mathcal{S}]} (\mu_+(\mathcal{T}) + \mu_-(\mathcal{T}))\right] =$$

$$= \mu(\mathcal{S}) - \sum_{\mathcal{T} \in \mathfrak{D}[\mathcal{S}]} \mu(\mathcal{T}) = 0.$$

Scanned with OKEN Scanner

Ca urmare, analog cu judecata obținerii expresiilor $W_{tot}[\mathcal{S}$ din paragrafele precedente, deducem că expresia costului total al clasificării lui $\mathcal{S}$ este

$$C_{tot}[\mathcal{S}] = \mu_+(\mathcal{S}) - \sum_{i=1}^{n} \mu_+(B_i) = \sum_{i=1}^{n} \mu_-(B_i) - \mu_-(\mathcal{S}) \qquad (9.2.8)$$

observînd că dacă μ, μ_+ și μ_- sînt cele din propoziția 9.2.1, atunci costul total al clasificării este

$$C_{tot}[\mathcal{S}] = W_{tot}[\mathcal{S}] \qquad (9.2.9)$$

căci $W_{tot}[B_i] = H(B_i) - H(B_i) = 0$, ceea ce dă o altă interpretare mărimii $W_{tot}[\mathcal{S}]$ pe care, pînă acum, am numit-o, după caz, corelație totală sau organizare totală. Într-adevăr, în baza relației (9.2.9) putem spune că $W_{tot}[\mathcal{S}]$ reprezintă costul total al clasificării elementelor lui $\mathcal{S}$ sau pierderea totală de coeziune sau de organizare atunci cînd $\mathcal{S}$ este descompus în părțile sale componente. Dacă unei desfaceri $\mathcal{D}[\mathcal{S}]$ i s-a asociat pierderea totală de coeziune sau costul total al clasificării $C_{tot}[\mathcal{S}]$, atunci privind și analizînd procedeul invers, de constituire pas cu pas a lui $\mathcal{S}$, am putea spune că $C_{tot}[\mathcal{S}]$ poate reprezenta cîștigul total de coeziune sau de organizare la reconstituirea ansamblului $\mathcal{S}$.

Indiferent de situație, să remarcăm că dacă descompunerea lui $\mathcal{S}$ se desfășoară astfel încît după un număr de etape să se obțină desfacerea atomică a acestuia, atunci cum $C_{tot}[\mathcal{S}]$ este constant indiferent de ordinea în care se face descompunerea, putem clasifica elementele lui $\mathcal{S}$ în subclase arbitrare și acestea la rîndul lor în sub-subclase arbitrare pînă la obținerea blocurilor lui $\mathcal{S}$. Dacă însă scopul nu este acela al descompunerii finale atomice a lui $\mathcal{S}$, atunci, clasificarea lui $\mathcal{S}$ trebuie făcută astfel încît costul acesteia să fie minim. Definind costul unei ramificări sau desfaceri (9.2.5) sau (9.2.6) cu ajutorul expresiilor (9.2.2) — (9.2.4), putem obține efectiv un algoritm entropic de clasificare. Astfel, dacă $\mathcal{D}[\mathcal{S}] = \{\mathcal{S}_j\}_{j=1}^{m}$ este o primă desfacere a lui $\mathcal{S}$, adoptăm drept desfacere optimă $\mathcal{D}_0[\mathcal{S}]$ pe aceea pentru care

$$C[\mathcal{S}, \mathcal{D}_0[\mathcal{S}]] = \min_{\mathcal{D}} \{C[\mathcal{S}; \mathcal{D}[\mathcal{S}]]\} \qquad (9.2.10)$$

unde $C[\mathcal{S}; \mathcal{D}[\mathcal{S}]] = W[\mathcal{S}; \mathcal{D}[\mathcal{S}]]$, ceea ce ar constitui un prim pas al algoritmului. Relația (9.2.10) definind criteriul entropic de clasificare al costului minim sau al pierderii

Scanned with OKEN Scanner

minime de coeziune, devine regulă și pentru desfacerea subsistemelor lui $\mathfrak{D}_0[\mathcal{S}]$ pînă la desfacerea dorită a lui $\mathcal{S}$.

Dacă însă este vorba de constituirea lui $\mathcal{S}$ sau a unei părți a acestuia din elemente constitutive, relația (9.2.10) poate fi luată pentru cazul de maximizare a coeziunii dintre blocurile sau subsistemele a căror juxtapunere constituie sistemul $\mathcal{S}$.

Evident că în cazul în care n este mare, algoritmul este anevoios, ceea ce l-a determinat pe S. Watanabe (1969) să propună ca în locul unei desfaceri arbitrare $\mathfrak{D}[\mathcal{S}]$ să fie luate numai desfacerile cu cîte două elemente sau dichotomiile, relația (9.2.10) rămînînd în vigoare.

Pot fi date însă și alte criterii entropice de clasificare bazate pe legătura dintre diferite părți ale lui $\mathcal{S}$, fără considerarea costului clasificării. Fie $\mathfrak{D}[\mathcal{S}] = \{B, \mathcal{S} \setminus \{B\}\}$ și să considerăm unul din coeficienții de organizare $G[\mathcal{S}; \mathfrak{D}[\mathcal{S}]]$ definiți de relația (9.1.2). Să presupunem că $\mathcal{S} = \{B_i\}_{i \in I}$.

Definiția 9.2.2. Se numește lider al sistemului $\mathcal{S}$ un bloc B_{i_0} cu proprietatea

$$G[B_{i_0}, \mathcal{S} \setminus \{B_{i_0}\}] = \max_{i \in I} \{G[B_i, \mathcal{S} \setminus \{B_i\}]\} \qquad (9.2.11)$$

Definiția 9.2.3. Se numește subsistem lider sau lider multiplu al sistemului $\mathcal{S}$ un subsistem $\mathcal{T} = \{B_k\}_{k \in K \subset I}$ pentru care

$$1)\ G[B_k, \mathcal{S} \setminus \{B_k\}] = \max_{i \in I} \{G[B_i, \mathcal{S} \setminus \{B_i\}]\},\ (\forall)\ k \in K;$$

$$2)\ G[B_k, \mathcal{T} \setminus \{B_k\}] = G[B_l, \mathcal{T} \setminus \{B_l\}],\ (\forall)\ k \neq l \in K.$$

Definiția 9.2.4. Blocul B_{k_0} se numește lider absolut al sistemului $\mathcal{S}$ dacă

1) există $\mathcal{T} = \{B_k\}_{k \in K \subset I}$ cu $G[B_k, \mathcal{S} \setminus \{B_k\}] =$

$$= \max_{i \in I} \{G[B_i, \mathcal{S} \setminus \{B_i\}]\},\ (\forall)\ k \in K\ ;$$

2) B_{k_0} este lider al lui $\mathcal{T}$.

Următorul rezultat este imediat

Propoziția 9.2.2. În orice sistem

1) există cel puțin un lider și cel puțin un lider absolut;
2) orice lider absolut este lider al sistemului dar nu și reciproc.

Scanned with OKEN Scanner

În baza celor de mai sus putem defini drept criteriu de clasificare entropică criteriul liderului, care conduce la o desfacere finală a lui $\mathcal{S}$ care nu este totdeauna cea atomică.

Analog relațiilor (9.2.10) și (9.2.11) putem da un criteriu de clasificare entropică al maximei organizări,

$$G_{tot}[\mathcal{T}_0] = \max_{\mathfrak{D};\ \mathcal{T}\in\mathfrak{D}[\mathcal{S}]} \{G_{tot}[\mathcal{T}]\} \qquad (9.2.12)$$

unde maximul este luat după toate desfacerile posibile $\mathfrak{D}[\mathcal{S}]$ și după toate elementele fiecăreia dintre ele, procedeul aplicîndu-se pentru $\mathcal{S}$, $\mathcal{S}\setminus\{\mathcal{T}_0\}$ și așa mai departe, evident, fără a avea desfacerea finală atomică.

Definiția 9.2.5. Se spune că blocul B_{i_0} este indiferent față de subsistemul $\mathcal{S}\setminus\{B_{i_0}\}$ sau că B_{i_0} și $\mathcal{S}\setminus\{B_{i_0}\}$ se află într-o relație de indiferență dacă

$$G[B_{i_0}, \mathcal{S}\setminus\{B_{i_0}\}] = \min_{i\in I} \{G[B_i, \mathcal{S}\setminus\{B_i\}]\} \neq 0 \qquad (9.2.13)$$

și că B_{i_0} este izolat de restul sistemului dacă egalitatea este zero.

Analog definițiilor 9.2.3 și 9.2.4 putem introduce și noțiunea de subsistem indiferent și pe cea de indiferență absolută. Evident că în orice sistem $\mathcal{S}$ există cel puțin un element indiferent. În baza acestor concepte putem introduce un alt criteriu entropic de clasificare și anume criteriul elementului indiferent, remarcînd cu ușurință că, în cazul în care orice subsistem al lui $\mathcal{S}$ are o anumită organizare, criteriul indiferenței coincide cu criteriul costului minim pentru dichotomii al lui Watanabe.

Se pot introduce criterii entropice de clasificare bazate și pe măsurile unilaterale ale dependenței prezentate în paragraful 7.6.2. Astfel vom spune că $B_0 \in \mathcal{S}$ este lider unilateral al lui $\mathcal{S}$ dacă

$$U(\mathcal{S}\setminus\{B_0\},\ B_0) = \max_{B\in\mathcal{S}}\{U(\mathcal{S}\setminus\{B\},\ B)\} \qquad (9.2.14)$$

unde $U(\mathcal{T}_1, \mathcal{T}_2)$ este măsura dependenței unilaterale (7.6.38) și ca urmare putem clasifica elementele lui $\mathcal{S}$ astfel încît în clasament fiecare element este lider unilateral pentru subsistemul format din elementele care îl succed.

Rezultatele de mai sus pot fi generalizate prin utilizarea măsurilor entropice ponderate ale dependenței, prezentate în paragraful 7.6.3. Astfel, considerînd unul din coeficienții de corelație entropică ponderată (7.6.62) putem introduce noțiunea de lider ponderat al sistemului extinzînd definițiile

Scanned with OKEN Scanner

9.2.2.—9.2.4, precum și noțiunea de indiferență ponderată sau cea de lider unilateral ponderat și odată cu acestea se pot defini criterii corespunzătoare de clasificare entropică.

Presupunînd acum că S este mulțimea partenerilor unui joc cu strategii aleatoare, clasificarea entropică reprezintă de fapt coalizarea jucătorilor după un criteriu sau altul, cum ar fi criteriul de coalizare al costului minim sau al pierderii minime de coeziune definit de relația (9.2.10) și care poate fi generalizat prin considerarea utilităților acțiunilor partenerilor, caz în care devine criteriul de coalizare ponderată de cost minim.

10. Informație și probabilitate

10.1 Principiul informației maxime

Există suficiente evenimente aleatoare ale căror probabilități nu pot fi evaluate efectiv, deci practic nu poate fi cunoscută legea probabilistică după care se desfășoară experimentul ale cărui rezultate sînt aceste evenimente. Mai mult chiar există evenimente aleatoare care nu permit să se considere drept probabilități corespunzătoare frecvențele relative ce se obțin în urma unui set de observații. Într-o astfel de situație, conform principiului rațiunii insuficiente al lui P. S. Laplace (1749—1827), toate evenimentele trebuie considerate ca fiind egal probabile dacă nu există nici un fel de date sau de motive în baza cărora să poată fi considerate altfel.

Desigur că principiul se aplică atunci cînd este vorba de un număr finit n de evenimente. Adeseori, în urma unor observații efectuate asupra unui anumit experiment dispunem de un număr de valori medii ale unor variabile aleatoare care depind de rezultatele acestui experiment.

Evident că pentru o anumită valoare medie dată există mai multe repartiții probabilistice compatibile cu aceasta. Problema care se pune este de a determina pe cea mai bună dintre aceste repartiții compatibilă cu această valoare medie. Deci este necesar un criteriu sau un principiu de selecție a repartiției optime. Un astfel de criteriu îl constituie principiul informației maxime, în baza căruia trebuie determinată acea repartiție probabilistă care maximizează o anumită cantitate

Scanned with OKEN Scanner

de informație (directă sau condiționată), supusă unei mulțimi de restricții. Formal am putea scrie

$$[\max_{p(x)}] I[p(x)] \qquad \text{(a)}$$

$$1)\ \varphi_k(p(x)) = M_k,\ 1 \leqslant k \leqslant m \qquad \text{(b)} \qquad (10.1.1)$$

$$2)\ \sum_x p(x) = 1 \text{ sau } \int_R p(x)\,dx = 1,\ p(x) \geqslant 0, \text{ (c)}$$

unde $I[p(x)]$ este cantitatea de informație corespunzătoare experimentului în studiu, $p(x)$ legea probabilistă ce descrie experimentul și pe care o căutăm, iar $\varphi_k(p(x)) = M_k$, $1 \leqslant k \leqslant m$, mulțimea celor m restricții la care repartiția este supusă, ca de exemplu valori medii. Relația (10.1.1) reprezintă de fapt o problemă de programare matematică, ce formalizează :

„Principiul informației maxime". Se alege drept lege probabilistă ce descrie desfășurarea unui experiment aleator, acea lege probabilistă care face maximă informația, sau nedeterminarea, compatibilă cu datele pe care le avem asupra experimentului aleator considerat.

Evident că dacă mulțimea (b) este mulțimea vidă, deci nu avem nici o restricție asupra repartiției căutate, atunci, problema (10.1.1) are ca soluție repartiția uniformă, ceea ce justifică faptul că principiul informației maxime extinde principiul rațiunii insuficiente al lui P. S. Laplace, sau că acesta este un caz particular al principiului informației maxime. Acest principiu a fost stabilit în mod independent de către S. Kullback și R. A. Leibler în 1951, E. T. Jaynes în 1957 și R. S. Ingarden în 1963. Unele rezultate în acest sens a avut și J. Kampé de Fériet în 1963 și din nou R. S. Ingarden împreună cu A. Kossakowski în 1971. Pentru faptul că acest principiu a fost aplicat pentru prima dată de către E. T. Jaynes în 1957 în mecanica statistică cuantică, permițîndu-i obținerea rapidă și elegantă a repartițiilor aleatoare de bază care intervin în acest domeniu și fundamentarea termodinamicii stărilor de echilibru, ceea ce a constituit de altfel piatra de încercare a acestuia, principiul informației maxime este cunoscut îndeosebi în asociere cu numele lui E. T. Jaynes, deci ca principiul informației maxime al lui E. T. Jaynes.

În practică, principiul informației maxime se dovedește a fi de fapt un algoritm de determinare a repartiției proba-

Scanned with OKEN Scanner

biliste care face maximă entropia, compatibilă cu datele pe care le avem la dispoziţie. El a putut fi formulat după ce deja se cunoştea entropia ca măsură a informaţiei şi constituie unul din primele rezultate fundamentale apărute în primul deceniu de la apariţia teoriei informaţiei.

Fie acum $\{\Omega, \mathfrak{K}, \nu\}$ un cîmp de probabilitate şi f şi g două variabile aleatoare mărginite. Să notăm cu

$$\Phi(\beta) = \int_{\Omega} e^{-\beta f(\omega)} d\nu(\omega) \tag{10.1.2}$$

unde β este o variabilă reală şi cu

$$G(\beta) = \int_{\Omega} g(\omega) e^{\beta g(\omega)} d\nu(\omega). \tag{10.1.3}$$

Propoziţia 10.1.1. Dacă (S. Guiaşu (1968, 1977))

$$\nu(\{\omega/g(\omega) > 0\}) > 0 \quad \text{şi} \quad \nu(\{\omega/g(\omega) < 0\}) > 0 \tag{10.1.4}$$

atunci funcţia $G(\beta)$ este crescătoare şi ecuaţia

$$G(\beta) = 0 \tag{10.1.5}$$

are soluţie unică.

Demonstraţie. Avem

$$G'(\beta) = \int_{\Omega} g^2(\omega) e^{\beta g(\omega)} d\nu(\omega) > 0$$

deci $G(\beta)$ este crescătoare pe $(-\infty, \infty)$. Să notăm acum cu

$$A = \{\omega/g(\omega) > 0\} \text{ şi } B = \{\omega/g(\omega) \leqslant 0\}.$$

Ca urmare putem scrie că

$$\lim_{\beta \to \infty} G(\beta) = \lim_{\beta \to \infty} \int_{A} g(\omega) e^{\beta g(\omega)} d\nu(\omega) + \lim_{\beta \to \infty} \int_{B} g(\omega) e^{\beta g(\omega)} d\nu(\omega) =$$

$$= \infty + 0 = \infty$$

Scanned with OKEN Scanner

precum și că

$$\lim_{\beta\to-\infty} G(\beta) = \lim_{\beta\to-\infty} \int_A g(\omega)e^{\beta g(\omega)}d\nu(\omega) + \lim_{\beta\to-\infty} \int_B g(\omega)e^{\beta g(\omega)}d\nu(\omega) =$$

$$= 0 - \infty = -\infty.$$

Deoarece $G(\beta)$ este crescătoare pe $(-\infty, \infty)$ și $\lim_{\beta\to\pm\infty} G(\beta) = \pm\infty$ rezultă că ecuația (10.1.5) are soluție unică.

Definiția 10.1.1. Spunem că variabila aleatoare mărginită f este nedegenerată dacă aproape peste tot

$$f \neq M(f) \tag{10.1.6}$$

unde $M(f)$ este valoarea medie a acesteia

$$M(f) = \int_\Omega f(\omega)d\nu(\omega). \tag{10.1.7}$$

Propoziția 10.1.2. Dacă f este o variabilă aleatoare mărginită nedegenerată atunci ecuația

$$\frac{\Phi'(\beta)}{\Phi(\beta)} = -M(f) \tag{10.1.8}$$

are soluție unică.

Demonstrație. Din (10.1.2) și (10.1.8) deducem că

$$\int_\Omega [M(f) - f(\omega)]e^{-\beta f(\omega)}d\nu(\omega) = 0$$

și multiplicînd cu constanta $e^{\beta M(f)}$ obținem

$$\int_\Omega [M(f) - f(\omega)]\, e^{\beta[M(f)-f(\omega)]}d\nu(\omega) = 0. \tag{10.1.9}$$

Ținînd cont că f este nedegenerată și notînd cu

$$g(\omega) = M(f) - f(\omega) \tag{10.1.10}$$

Scanned with OKEN Scanner

ecuaţia (10.1.9) coincide cu ecuaţia (10.1.5) şi ca urmare propoziţia este demonstrată.

Propoziţia 10.1.3. Repartiţia aleatoare

$$p_k \geqslant 0,\ 1 \leqslant k \leqslant n,\ \sum_{k=1}^{n} p_k = 1 \qquad (10.1.11)$$

care maximizează entropia corespunzătoare

$$H_n = H_n[p_1, \ldots, p_n] = -\sum_{k=1}^{n} p_k \log_e p_k \qquad (10.1.12)$$

dacă valoarea medie a variabilei f este cunoscută

$$M(f) = \sum_{k=1}^{n} f_k p_k \qquad (10.1.13)$$

are forma

$$p_k = \frac{1}{\Phi(\beta)} e^{-\beta f_k},\ 1 \leqslant k \leqslant n, \qquad (10.1.14)$$

unde

$$\Phi(\beta) = \sum_{k=1}^{n} e^{-\beta f_k} \qquad (10.1.15)$$

β fiind soluţia unică a ecuaţiei (10.1.8), caz în care informaţia maximă este

$$H_n^{\max} = \log_e \Phi(\beta) + \beta M(f). \qquad (10.1.16)$$

Demonstraţie. Rezultatul se obţine imediat aplicînd metoda multiplicatorilor lui Lagrange pentru funcţia (10.1.12) cu condiţiile (10.1.11) şi (10.1.13).

În relaţia (10.1.12) sînt consideraţi logaritmii naturali pentru comoditate în scriere.

Să presupunem că variabila aleatoare f reprezintă utilitatea sau cîştigul pe care un jucător J îl are într-o partidă prin aplicarea unei anumite strategii aleatoare, ca în paragraful 8.1.2. Atunci valoarea medie a variabilei f reprezintă utilitatea medie sau cîştigul mediu corespunzător strategiei aplicate. În acest context, propoziţia 10.1.3 determină strategia optimă (10.1.14) ce trebuie adoptată pentru a maximiza cantitatea de informaţie (10.1.12), ştiind cît este sau

Scanned with OKEN Scanner

cît trebuie să fie utilitatea medie a jocului, ea însăşi compatibilă cu jocul celorlalţi parteneri.

Putem extinde acum propoziţia 10.1.3 la cazul în care în locul informaţiei (10.1.12) considerăm entropia ponderată.

Propoziţia 10.1.4. Repartiţia aleatoare (10.1.11) care maximizează entropia ponderată corespunzătoare

$$\Gamma_n[u, p] = -\sum_{k=1}^{n} u_k p_k \log p_k, \quad u_k \geqslant 0, \ 1 \leqslant k \leqslant n \qquad (10.1.17)$$

cu condiţia (10.1.13) are forma

$$p_k = \mathrm{e}^{-1-\frac{\alpha+\beta f_k}{u_k}}, \quad 1 \leqslant k \leqslant n \qquad (10.1.18)$$

α şi β fiind soluţii ale ecuaţiilor

$$\sum_{k=1}^{n} \mathrm{e}^{-\frac{\alpha+\beta f_k}{u_k}} = \mathrm{e}; \quad \sum_{k=1}^{n} f_k \mathrm{e}^{-\frac{\alpha+\beta f_k}{u_k}} = \mathrm{e}M(f). \qquad (10.1.19)$$

Demonstraţie. Rezultatul se obţine de asemenea aplicînd metoda multiplicatorilor lui Lagrange, fără a mai avea asigurată însă unicitatea soluţiei pentru parametrii α şi β.

Dacă $f_k = u_k$, $1 \leqslant k \leqslant n$, putînd avea deci drept valoare medie utilitatea medie, dar nu neapărat, atunci

$$p_k = \frac{\mathrm{e}^{-\frac{\alpha}{u_k}}}{\sum_{k=1}^{n} \mathrm{e}^{-\frac{\alpha}{u_k}}} \quad 1 \leqslant k \leqslant n \qquad (10.1.20)$$

parametrul α fiind soluţie a ecuaţiei

$$\sum_{k=1}^{n} u_k \mathrm{e}^{-\frac{\alpha}{u_k}} = M(f) \sum_{k=1}^{n} \mathrm{e}^{-\frac{\alpha}{u_k}} \qquad (10.1.21)$$

caz în care

$$H_n^{\max} = \log_{\mathrm{e}} \sum_{k=1}^{n} \mathrm{e}^{-\frac{\alpha}{u_k}} + \frac{\alpha}{\sum_{k=1}^{n} \mathrm{e}^{-\frac{\alpha}{u_k}}} \sum_{k=1}^{n} \frac{1}{u_k} \mathrm{e}^{-\frac{\alpha}{u_k}} \qquad (10.1.22)$$

Scanned with OKEN Scanner

Revenind acum la propoziția (10.1.3 putem deduce diverse rezultate particulare corespunzînd unor repartiții statistice clasice.

Propoziția 10.1.5. Repartiția continuă

$$p(x) \geqslant 0, \int_R p(x)\,dx = 1 \tag{10.1.23}$$

care maximizează entropia corespunzătoare

$$H = H(p) - \int_R p(x) \log_e p(x)\,dx \tag{10.1.24}$$

compatibilă cu condițiile

$$M(X) = m = \int_R xp(x)\,dx \text{ și } M(X^2) = \sigma^2 - m^2 = \int_R x^2 p(x)\,dx \tag{10.1.25}$$

este repartiția normală

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-m)^2}{2\sigma^2}} \tag{10.1.26}$$

caz în care

$$H^{max} = \log_e \sigma\sqrt{2\pi e}. \tag{10.1.27}$$

Demonstrație. Considerînd multiplicatorii lui Lagrange α, β și γ și funcția

$$F(p(x), \alpha, \beta, \gamma) = H(p) -$$

$$- \alpha\int_R p(x)\,dx - \beta\int_R xp(x)\,dx - \gamma\int_R x^2 p(x)\,dx$$

avem

$$F(p(x), \alpha, \beta, \gamma) = -\int_R p(x) \log_e p(x)\, e^{\alpha+\beta x+\gamma x^2}\,dx \leqslant$$

$$\leqslant \int_R p(x)\left(\frac{e^{-\alpha-\beta x-\gamma x^2}}{p(x)} - 1\right)dx$$

Scanned with OKEN Scanner

cu egalitate dacă și numai dacă

$$p(x) = e^{-\alpha-\beta x-\gamma x^2}. \tag{10.1.28}$$

Impunînd condițiile (10.1.23) și (10.1.25) obținem sistemul neliniar

$$\begin{cases} \sqrt{\gamma}\, e^{\alpha-\frac{\beta^2}{4\gamma}} = \sqrt{\pi} \\ -\beta\sqrt{\pi}\, e^{\frac{\beta^2}{4\gamma}-\alpha} = 2m \\ \sqrt{\pi}(\beta^2+2\gamma)\, e^{\frac{\beta^2}{4\gamma}-\alpha} = 4\gamma^2\sqrt{\gamma}\,(\sigma^2-m^2) \end{cases} \tag{10.1.29}$$

a cărui soluționare identifică relațiile (10.1.28) și (10.1.26). Printr-un calcul simplu deducem și relația (10.1.27).

În cazul particular $m = 0$, propoziția 10.1.5 a fost demonstrată, ținînd cont de propozițiile 10.1.2 și 10.1.3, de către J. Kampé de Fériet (1963).

Să presupunem acum că $\Omega = \{\omega_0, \omega_1, \ldots, \omega_n, \ldots\}$ este o mulțime numărabilă, $\mathcal{K}$ o clasă de submulțimi ale lui Ω și ν o probabilitate pe $\mathcal{K}$. Să notăm cu $\nu_n = \nu(\omega_n)$, $n = 0, 1, 2, \ldots$

Să înlocuim probabilitatea inițială ν printr-o nouă probabilitate p, absolut continuă în raport cu ν și să notăm cu $p_n = p(\omega_n)$, $n = 0, 1, 2, \ldots$

Variația cantității de informație corespunzătoare acestei înlocuiri va fi

$$H(p/\nu) = \sum_{n=0}^{\infty} p_n \log_e \frac{p_n}{\nu_n}. \tag{10.1.30}$$

Presupunînd

$$\nu_n = \frac{e^{-1}}{n!}, \quad n = 0, 1, 2, \ldots \tag{10.1.31}$$

atunci

$$H(p/\nu) = 1 + \sum_{n=0}^{\infty} p_n \log_e (p_n \cdot n!). \tag{10.1.32}$$

Să presupunem că $f_n = f(\omega_n) = n$ și ca urmare

$$M(f) = \sum_{n=0}^{\infty} n\, p_n = \theta. \tag{10.1.33}$$

Scanned with OKEN Scanner

Cu aceste precizări putem formula următoarea propoziție datorată lui R. S. Ingarden și A. Kossakowski (1971).

Propoziția 10.1.6. Repartiția aleatoare

$$p_n \geqslant 0, \sum_{n=0}^{\infty} p_n = 1 \tag{10.1.34}$$

care maximizează expresia

$$H = 1 - H(p/\nu) \tag{10.1.35}$$

cu condiția (10.1.33) este tocmai repartiția Poisson

$$p_n = \theta^n \frac{e^{-\theta}}{n!}, \quad n = 0, 1, 2, \ldots \tag{10.1.36}$$

Demonstrație. Utilizînd multiplicatorii lui Lagrange α și β avem

$$F(p_n, \alpha, \beta) = H - \alpha \sum_{n=0}^{\infty} p_n - \beta \sum_{n=0}^{\infty} n\, p_n \leqslant$$

$$\leqslant \sum_{n=0}^{\infty} p_n \left[\frac{e^{-\alpha-\beta n}}{n!\, p_n} - 1 \right]$$

cu egalitate dacă și numai dacă

$$p_n = \frac{e^{-\alpha-\beta n}}{n!}, \quad n = 0, 1, 2, \ldots \tag{10.1.37}$$

Impunînd condițiile (10.1.33) și (10.1.34) găsim $\alpha = \theta = e^{-\beta}$ și ca urmare relațiile (10.1.37) și (10.1.36) coincid.

Propoziția 10.1.7. Repartiția aleatoare continuă

$$p(x) \geqslant 0, \int_0^{\infty} p(x)\, dx = 1 \tag{10.1.38}$$

care maximizează entropia corespunzătoare

$$H = H(p) = -\int_0^{\infty} p(x) \log_e p(x)\, dx \tag{10.1.39}$$

Scanned with OKEN Scanner

compatibilă cu condiţia

$$M(X) = m = \int_0^\infty xp(x)\,dx \qquad (10.1.40)$$

este repartiţia exponenţială

$$p(x) = \frac{1}{m}\, e^{-\frac{x}{m}};\ x \in [0, \infty),\ m > 0, \qquad (10.1.41)$$

caz în care

$$H^{\max} = \log_e me \qquad (10.1.42)$$

Demonstraţie. Considerînd multiplicatorii lui Lagrange avem $F(p, \alpha, \beta) = H(p) - \alpha \int_0^\infty p(x)\,dx - \beta \int_0^\infty xp(x)\,dx =$

$$= \int_0^\infty p(x) \log_e \frac{e^{-\alpha-\beta x}}{p(x)}\,dx \leqslant \int_0^\infty p(x)\left[\frac{e^{-\alpha-\beta x}}{p(x)} - 1\right]dx \text{ cu egalitate}$$

dacă şi numai dacă $p(x) = e^{-\alpha-\beta x}$.

Impunînd condiţiile (10.1.38) şi (10.1.40) deducem $e^{-\alpha} = \beta$ şi $\beta = \frac{1}{m}$ de unde (10.1.41) şi ca urmare (10.1.42).

Propoziţia 10.1.8. Repartiţia aleatoare discretă

$$p(x) \geqslant 0,\ \sum_{x \in N} p(x) = 1 \qquad (10.1.43)$$

care maximizează entropia corespunzătoare

$$H = H(p) = -\sum_{x \in N} p(x) \log_e p(x) \qquad (10.1.44)$$

compatibilă cu condiţia

$$M(X) = \sum_{x \in N} xp(x) = m > 0 \qquad (10.1.45)$$

este repartiţia geometrică

$$p(x) = \frac{1}{1+m}\left(\frac{m}{1+m}\right)^x,\ x \in N \qquad (10.1.46)$$

Scanned with OKEN Scanner

caz în care

$$H^{\max} = \log_e(m+1) + \frac{1}{m}\log_e \frac{m}{1+m}. \qquad (10.1.47)$$

Demonstrație. Procedînd ca mai sus avem

$$F(p,\ \alpha,\ \beta) = \sum_{x\in N} p(x)\log_e \frac{e^{-\alpha-\beta x}}{p(x)} \leqslant \sum_{x\in N} p(x)\left[\frac{e^{-\alpha-\beta x}}{p(x)} - 1\right] \text{ cu}$$

egalitate dacă și numai dacă $p(x) = e^{-\alpha-\beta x}$, funcție pentru care, impunînd condițiile (10.1.43) și (10.1.45) deducem $e^{-\alpha} = 1 - e^{-\beta}$ și $m(e^{\beta} - 1) = 1$ de unde rezultă (10.1.46) și (10.1.47).

În funcție de condițiile în raport cu care trebuie să fie compatibilă maximizarea entropiei Shannon, putem obține și alte repartiții clasice. Spre exemplu, dacă în propoziția 10.1.7 adăugăm și condiția

$$\int_0^{\infty} p(x)\log_e x\,dx = A \qquad (10.1.48)$$

atunci vom obține repartiția gama generalizată, din care, în particular putem deduce repartiția gama simplă și alte repartiții înrudite cu aceasta.

Să remarcăm, ținînd cont de cele de mai sus, că dacă în practică nu dispunem de multe date experimentale pentru evaluarea probabilităților prin procedeul clasic al statisticii matematice, selecție — estimare — verificări de ipoteze, dar avem totuși anumite date, atunci putem determina repartiția aleatoare ca fiind aceea care face maximă nedeterminarea, compatibilă cu aceste cîteva date. Această repartiție se bucură de proprietatea de a furniza cea mai mare cantitate de informație compatibilă cu datele de care dispunem, astfel încît noțiunea de informație și măsura cantității de informație permit determinarea efectivă a repartiției probabiliste căutate.

10.2. Conexiunea dintre informație și probabilitate

Rod al preocupărilor multor oameni de știință din domenii de cercetare diferite, teoria informației apare în cele din urmă ca fiind creația lui Claude E. Shannon care, în lucrarea sa fundamentală din 1948 a dat formula entropiei, ca măsură

Scanned with OKEN Scanner

a gradului de nedeterminare pe care îl conţine un anumit experiment aleator, sau a cantităţii de informaţie pe care realizarea acestuia o furnizează experimentatorului. Conceptul de entropie şi măsura lui Shannon pentru acesta, au determinat alte preocupări ştiinţifice cu deosebite rezultate atît privind rigurozitatea cît şi aplicabilitatea ideilor lui Shannon. Dacă $\{\Omega, \mathcal{K}, P\}$ este un cîmp de probabilitate, unde Ω este mulţimea evenimentelor elementare, $\mathcal{K}$ este o σ-algebră de părţi ale mulţimii Ω, iar P o probabilitate definită pe $\mathcal{K}$, satisfăcînd axiomatica lui N. A. Kolmogorov (1933),

1) $P(\Omega) = 1$;

2) $P(A) \geqslant 0$, $A \in \mathcal{K}$;

3) $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$, $A_i \in \mathcal{K}$, $A_i \cap A_j = \emptyset$, $i \neq j = 1, 2, \ldots$

atunci considerînd o desfacere a lui Ω, $\mathcal{D}[\Omega] = \{A_i\}_{i=1}^{m}$, $A_i \cap A_j = \emptyset$, $i \neq j$, $\bigcup_{i=1}^{n} A_i = \Omega$, entropia acestei partiţii în sensul lui C. E. Shannon este

$$H(\mathcal{D}(\Omega)) = -\sum_{i=1}^{n} P(A_i) \log_2 P(A_i). \qquad (10.2.1)$$

Toate consideraţiile şi rezultatele obţinute în teoria informaţiei s-au bazat pe formula (10.2.1). După cum se poate constata, pornind de la noţiunea de probabilitate, pe baza formulei (10.2.1) se obţine noţiunea de informaţie, ambele fiind concepte fundamentale. Această formulă ne dă cantitatea de informaţie pe care o conţine cîmpul finit de probabilitate generat de partiţia $\mathcal{D}(\Omega)$ a lui Ω, utilizînd probabilitatea drept concept primar. Se pune atunci în mod firesc întrebarea: „nu poate fi considerată noţiunea de informaţie ca o noţiune de mai mare generalitate decît noţiunea de probabilitate, sau cu alte cuvinte, nu putem considera conceptul de informaţie primar conceptului de probabilitate?" Răspunsul este evident afirmativ. Şi dacă este afirmativ, atunci apare în mod natural şi întrebarea inversă: „cum poate fi determinată probabilitatea pe baza informaţiei?"

Prima încercare de soluţionare a acestei probleme a fost efectuată de către R. S. Ingarden şi K. Urbanik (1962). Ei au formulat un sistem de axiome asupra informaţiei în baza cărora au dedus existenţa unei probabilităţi astfel încît legătura dintre conceptul de probabilitate, dedus, şi

Scanned with OKEN Scanner

conceptul de informaţie, primar, să fie asigurată tot de către entropia lui Shannon. Axiomatica lui Ingarden şi Urbanik este greoaie şi prin urmare nu s-a impus ca un rezultat hotărîtor în acest domeniu, chiar dacă ulterior s-a mai revenit asupra acesteia. A doua încercare de soluţionare a problemei construirii probabilităţii pe baza informaţiei aparţine lui J. Kampé de Fériet şi Bruno Forte (1967), care au ales o cale mai simplă de a lega cele două concepte. În loc să pornească de la axiomatizarea informaţiei pe care o conţinea un cîmp de evenimente, aşa cum făcuseră Ingarden şi Urbanik, J. K. de Fériet şi B. Forte au luat drept concept primar informaţia, pe care o conţine un singur eveniment. Acest model local de axiomatizare a informaţiei şi regăsire a probabilităţii a fost analizat şi simplificat de către S. Guiaşu (1968). Pornind de aici am formulat axiomele măsurării cantităţii de informaţie din paragraful 2.2. Ţinînd cont de acestea să demonstrăm

Propoziţia 10.2.1. Fie $\{\Omega, \mathcal{K}\}$ un cîmp de evenimente şi $I(A)$ funcţia ce satisface axiomele A.1 — A.6 din definiţia 2.2.1. Dacă funcţia din enunţul axiomei A.4 este continuă şi satisface condiţiile

$$f^{-1}: [0, \infty) \to [0, 1]; \quad f^{-1}(0) = 0 \text{ şi } f^{-1}(+\infty) = 0 \tag{10.2.2}$$

atunci funcţia de mulţime μ_f definită de egalitatea

$$\mu_f(A) = f^{-1}(I(A)) \quad (\forall) A \in \mathcal{K} \tag{10.2.3}$$

este o probabilitate definită pe σ-algebra evenimentelor $\mathcal{K}$.

Demonstraţie. Evident că

$$0 \leqslant \mu_f(A) \leqslant 1; \quad \mu_f(\emptyset) = 0; \quad \mu_f(\Omega) = 1 \quad (\forall) A \in \mathcal{K}. \tag{10.2.4}$$

Din egalităţile (2.2.3) şi (10.2.3) rezultă că

$$\mu_f(A \cup B) = \mu_f(A) + \mu_f(B) \quad (\forall) A, B \in \mathcal{K}, \; A \cap B = \emptyset \tag{10.2.5}$$

Pornind de la relaţia (10.2.5), din aproape în aproape deducem că μ_f este finit aditivă. Ţinînd cont şi de axioma A.5, rezultă că μ_f este o funcţie superior semicontinuă de mulţime, ca în relaţia (2.2.4) pentru $I(A)$ şi ca urmare, μ_f este o funcţie numărabil aditivă de mulţime.

Scanned with OKEN Scanner

Într-adevăr, fie $\{E_i\}_{i\in N}$, $E_i \in \mathcal{K}$, $E_i \cap E_j = \emptyset$, $i \neq j \in \mathbf{N}$ un șir de mulțimi și $E = \bigcup_{i\in N} E_i$. Notînd cu $A_n = \bigcup_{i=1}^{n} E_i$ obținem un șir de mulțimi $\{A_n\}_{n\in N}$ pentru care

a) $A_n \subset A_{n+1}$, $n \in \mathbf{N}$; b) $\lim_{n\to\infty} A_n = \bigcup_{n\in N} A_n = E$

Deoarece μ_f este superior semicontinuă rezultă că

$$\mu_f(\bigcup_{i\in N} E_i) = \mu_f(E) = \mu_f(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \mu_f(A_n) =$$

$$= \lim_{n\to\infty} \sum_{i=1}^{n} \mu_f(E_i) = \sum_{i=1}^{\infty} \mu_f(E_i)$$

și deci μ_f este probabilitate pe $\mathcal{K}$.

Propoziția 10.2.2. Fie $\{\Omega, \mathcal{K}\}$ un cîmp de evenimente și $I(A)$ funcția introdusă de definiția 2.2.1. Să presupunem că funcția f din enunțul axiomei $A.4$ este continuă și satisface condițiile (10.2.2) și fie μ_f probabilitatea care corespunde lui $I(A)$ conform propoziției 10.2.1. În aceste condiții dacă independența informațională a evenimentelor din $\mathcal{K}$ coincide cu independența probabilistă ($I(A \cap B) = I(A) + I(B) \Leftrightarrow$ $\Leftrightarrow \mu_f(A \cap B) = \mu_f(A)\cdot\mu_f(B)$) atunci

$$f(x) = c \log \frac{1}{x},\ c > 0; \qquad (10.2.6)$$

și

$$\mu_f(A) = e^{-\frac{1}{c} I(A)} \text{ sau } I(A) = -c \log \mu_f(A), c > 0 \qquad (10.2.7)$$

Demonstrație. Conform ipotezei, $I(B_1 \cap B_2) = I(B_1) + I(B_2) \Leftrightarrow \mu_f(B_1 \cap B_2) = \mu_f(B_1)\cdot\mu_f(B_2)$ pentru orice B_1 și $B_2 \in \mathcal{K}$. Să punem $x_1 = I(B_1)$ și $x_2 = I(B_2)$ și ca urmare $f^{-1}(x_1 + x_2) = f^{-1}(I(B_1) + I(B_2)) = f^{-1}(I(B_1 \cap B_2)) = \mu_f(B_1 \cap B_2) = \mu_f(B_1)\cdot\mu_f(B_2) = f^{-1}(I(B_1))\cdot f^{-1}(I(B_2)) = f^{-1}(x_1) f^{-1}(x_2)$.

Dar ecuația funcțională $f^{-1}(x_1 + x_2) = f^{-1}(x_1) f^{-1}(x_2)$, pentru $\varphi = \log f^{-1}$ sau $f^{-1} = e^{\varphi}$ este echivalentă cu ecuația funcțională a lui Cauchy $\varphi(x_1 + x_2) = \varphi(x_1) + \varphi(x_2)$ a cărei soluție unică este $\varphi = cx$. Prin urmare, $f^{-1}(x) = e^{-\frac{x}{c}}$, $c > 0$ și deci $f(x) = c \log \frac{1}{x}$, $c > 0$, observînd că măsura corespunzătoare μ_f este o probabilitate întrucît $f^{-1}(0) = 1$ și $f^{-1}(+\infty) = 0$ și demonstrația este încheiată.

Ca urmare a acestor rezultate plecînd de la noţiunea de probabilitate, formula

$$I(A) = -\log P(A) \tag{10.2.8}$$

este doar una din formele conexiunii dintre informaţie şi probabilitate.

Permiţînd introducerea unei probabilităţi pe baza informaţiei, axiomele A.1—A.6 din definiţia 2.2.1 reliefează un alt aspect al conexiunii dintre cele două concepte, fiecare din ele fiind pe rînd concept primar sau concept derivat din conceptul primar. Practic, informaţia serveşte efectiv la construirea probabilităţii atunci cînd acest lucru este cerut, prin utilizarea principiului informaţiei maxime. Există oare şi alte posibilităţi de a determina probabilitatea pornind de la informaţie? Presupunînd că informaţia poate fi măsurată, răspunsul este afirmativ, în baza relaţiei (10.2.7).

Să considerăm acum un anumit experiment aleator oarecare descris de o variabilă aleatoare X pe care, ţinînd cont de cele de mai sus, o putem scrie ca

$$X : \begin{pmatrix} x \\ p(x) \end{pmatrix} \text{ sau } X : \begin{pmatrix} x \\ I(x) \end{pmatrix} \tag{10.2.9}$$

unde $p(x)$ este probabilitatea evenimentului $\{X = x\}$ în cazul discret, iar $I(x)$ cantitatea de informaţie furnizată de realizarea acestuia. Acestea fiind legate prin relaţia (10.2.7), rezultă echivalenţa

$$\sum_x p(x)=1 \Leftrightarrow \sum_x \mathrm{e}^{-I(x)}=1 \text{ sau } \int_R p(x)\,\mathrm{d}x=1 \Leftrightarrow \int_R \mathrm{e}^{-I(x)}\,\mathrm{d}x = 1 \tag{10.2.10}$$

după cum x este discretă sau continuă, în $I(x)$ fiind incorporată şi constanta c din relaţia (10.2.7). Aşadar, dacă în urma efectuării unui experiment am putea măsura cantitatea de informaţie pe care acesta o furnizează în fiecare din evenimentele ce îl definesc, atunci am putea deduce şi repartiţia probabilistă corespunzătoare acestuia.

Propoziţia 10.2.3. Fie un experiment aleator cu n rezultate posibile $x_1, x_2, \ldots, x_n$. Dacă realizarea oricăruia dintre acestea furnizează aceeaşi cantitate de informaţie, deci $I(x_i) = \alpha$, $1 \leqslant i \leqslant n$, atunci repartiţia probabilistă corespunzătoare acestui tip de experiment este repartiţia uniformă discretă.

Scanned with OKEN Scanner

Demonstrație. Într-adevăr, dacă $I(x_i) = \alpha$, $1 \leqslant i \leqslant n$, atunci din (10.2.10) rezultă $\sum_{i=1}^{n} e^{-I(x_i)} = n e^{-\alpha} = 1 \Rightarrow \alpha = \ln n$ și din (10.2.7) rezultă că $p(x_i) = \frac{1}{n}$, $1 \leqslant i \leqslant n$.

Propoziția 10.2.4. Fie un experiment cu două rezultate posibile A și $\bar{A}$ a căror realizare furnizează respectiv cantitatea de informație $I(A) = \alpha$ și $I(\bar{A}) = \beta$. Să presupunem că efectuăm acest experiment în condiții identice de x ori astfel că

1) rezultatul fiecărei probe este independent de precedenta probă.

2) rezultatul A apare în cea de a x-aprobă.

Atunci cantitatea de informație corespunzătoare celor x probe este

$$I(x) = \alpha + \beta(x - 1), \qquad (10.2.11)$$

iar repartiția probabilistă corespunzătoare experimentului este repartiția geometrică

$$p(x) = pq^{x-1}, \quad x = 1, 2, \ldots \qquad (10.2.12)$$

unde $p = P(A)$ și $q = P(\bar{A}) = 1 - p$.

Demonstrație. Relația (10.2.11) apare ca urmare a axiomei independenței A.6: Din (10.2.10) rezultă că

$$1 = \sum_{x=1}^{\infty} e^{-I(x)} = e^{\beta-\alpha} \sum_{x=1}^{\infty} [e^{-\beta}]^x = \frac{e^{-\alpha}}{1 - e^{-\beta}}$$

și din (10.2.7) rezultă că

$$p(x) = (1 - e^{-\beta})\, e^{-\beta(x-1)}, \quad x = 1, 2, \ldots$$

Notînd $1 - e^{-\beta} = p$ și $e^{-\beta} = q$ rezultă relația (10.2.12).

Consecință. Dacă $\alpha = \beta$, adică informația de realizare a rezultatului A este egală cu cea de nerealizare, atunci $I(x) = \alpha x$ și $p(x) = \frac{1}{2^x}$, $x = 1, 2, \ldots$

Propoziția 10.2.5. Fie un experiment cu două rezultate posibile A și $\bar{A}$ a căror realizare furnizează respectiv cantitatea

Scanned with OKEN Scanner

de informație $I(A) = \alpha$ și $I(\bar{A}) = \beta$. Să presupunem că experimentul se repetă de n ori în condiții identice astfel că

1) rezultatul fiecărei probe nu depinde de precedenta probă

2) A apare de x ori și $\bar{A}$ de $(n - x)$ ori.

Atunci cantitatea de informație corespunzătoare celor n probe cu x realizări este

$$I(n\,;\,x) = \alpha x + \beta(n - x), \tag{10.2.13}$$

iar repartiția probabilistă corespunzătoare acestui experiment este repartiția binomială

$$P(n\,;\,x) = C_n^x\, p^x\, q^{n-x}. \tag{10.2.14}$$

unde $p = P(A)$ și $q = P(\bar{A}) = 1 - p$

Demonstrație. Dacă în n probe cu x realizări $I(n\,;\,x) = \alpha x + \beta(n - x)$, atunci, cum un astfel de experiment poate conduce la C_n^x situații ce pot da aceeași cantitate de informație $I(n\,;\,x)$, conform cu (10.2.10) avem $1 = \sum_{x=1}^{n} C_n^x\, e^{-I(n\,;\,x)} = e^{-\beta n}(1 + e^{\beta-\alpha})^n \Rightarrow e^{\beta} = 1 + e^{\beta-\alpha}$ și ținînd cont și de (10.2.7) deducem că pentru un grup de n probe cu x realizări

$$P_1(n : x) = (1 - e^{-\beta})^x\, e^{-\beta(n-x)} \tag{10.2.15}$$

și ca urmare, pentru cele C_n^x situații identice informațional rezultă

$$P(n\,;\,x) = C_n^x\, P_1(n\,;\,x) \tag{10.2.16}$$

relație din care, notînd $p = 1 - q = e^{-\beta}$, deducem (10.2.14). În mod analog, în funcție de forma expresiei $I(x)$, putem obține și alte repartiții clasice.

Propoziția 10.2.6. Să presupunem acum un experiment aleator în care

$$I(x) = \alpha x + \beta \ln x! + \gamma,\ \alpha, \beta, \gamma \geqslant 0,\ x \in \mathbf{N} \tag{10.2.17}$$

Atunci

$$p(x) = A\frac{e^{-\alpha x}}{(x!)^{\beta}};\qquad A^{-1} = \sum_{x=0}^{\infty} e^{-\alpha x}\,(x!)^{-\beta}. \tag{10.2.18}$$

Scanned with OKEN Scanner

Demonstrație. Într-adevăr $1 = \sum_{x=0}^{\infty} e^{-I(x)} = e^{-\gamma} \sum_{x=0}^{\infty} e^{-\alpha x}(x!)^{-\beta} \Rightarrow$

$\Rightarrow \sum_{x=0}^{\infty} \frac{e^{-\alpha x}}{(x!)^{\beta}} = e^{\gamma}$ și ținînd cont de (10.2.7) obținem relația (10.2.18).

Consecință. Dacă

1) $\beta = 0$ atunci $I(x) = \alpha x + \gamma$ și $e^{\gamma} = \frac{1}{1 - e^{-\alpha}} \Rightarrow p(x) = (1 - e^{-\alpha})\, e^{-\alpha x}$, adică repartiția geometrică;

2) $\beta = 1$ atunci (10.2.18) reprezintă repartiția Poisson de parametru $e^{-\alpha}$;

3) $\alpha = 0$ atunci $p(x) = \frac{e^{-\gamma}}{(x!)^{\beta}}$, $x \in \mathbf{N}$; $e^{\gamma} = \sum_{x=0}^{\infty} (x!)^{-\beta}$;

4) $\alpha = 0$ și $\beta = 1$ atunci $p(x) = \frac{e^{-1}}{x!}$, $x \in \mathbf{N}$.

Propoziția 10.2.7. Dacă $I(x) = \alpha$ pentru orice $x \in [a, b]$ atunci avem repartiția uniformă pe $[a, b]$ adică $p(x) = \frac{1}{b - a}$.

Demonstrație. Într-adevăr, deoarece $1 = \int_a^b e^{-I(x)}\, dx =$

$= e^{-\alpha}(b - a) \Rightarrow \alpha = \ln(b - a)$ și din (10.2.7) rezultă afirmația făcută.

Propoziția 10.2.8. Dacă experimentul este definit astfel încît cantitatea de informație depinde liniar de rezultat, adică

$$I(x) = \alpha x + \beta, \quad x \in [0, \infty), \quad \alpha, \beta \geqslant 0 \qquad (10.2.19)$$

atunci experimentul este descris de repartiția exponențială de parametru α.

Demonstrație. Într-adevăr

$1 = \int_0^{\infty} e^{-\alpha x - \beta}\, dx = e^{-\beta} \frac{1}{\alpha} \Rightarrow \alpha = e^{-\beta} \Rightarrow p(x) = \alpha e^{-\alpha x}$, adică repartiția exponențială.

Extinzînd propoziția 10.2.8 putem formula

Propoziția 10.2.9. dacă $I(x) = \alpha |x| + \beta$, $x \in \mathbf{R}$, $\alpha, \beta \geqslant 0$, atunci experimentul este descris de repartiția Laplace

$$p(x) = \frac{\alpha}{2} e^{-\alpha |x|}, \quad x \in \mathbf{R} \qquad (10.2.20)$$

Scanned with OKEN Scanner

Demonstrație. Într-adevăr,

$$1 = e^{-\beta}\int_{\mathbf{R}} e^{-\alpha|x|}\,dx = \frac{2e^{-\beta}}{\alpha} \Rightarrow e^{-\beta} = \frac{\alpha}{2}$$ și din (10.2.7) rezultă (10.2.20).

Propoziția 10.2.10. Dacă experimentul este definit astfel încît cantitatea de informație depinde parabolic de rezultat, adică

$$I(x) = \alpha x^2 + \beta x + \gamma,\ x \in \mathbf{R},\ \alpha > 0,\ \beta^2 - 4\alpha\gamma \leqslant 0,\ \beta, \gamma \in \mathbf{R} \tag{10.2.21}$$

atunci repartiția probabilistă corespunzătoare acestuia este

$$p(x) = \sqrt{\frac{\alpha}{\pi}}\, e^{-\alpha\left(x + \frac{\beta}{2\alpha}\right)^2} \tag{10.2.22}$$

Demonstrație. Într-adevăr

$$1 = e^{\frac{\beta^2 - 4\gamma}{4\alpha}}\int_{\mathbf{R}} e^{-\alpha\left(x + \frac{\beta}{2\alpha}\right)^2} dx \Rightarrow e^{\frac{\beta^2 - 4\alpha\gamma}{4\alpha}} = \sqrt{\frac{\alpha}{\pi}}$$ și ținînd cont de (10.2.7) rezultă (10.2.22).

Consecință. Dacă

1) $\alpha = \frac{1}{2\sigma^2}$ și $\beta = -\frac{m}{\sigma^2}$ atunci (10.2.22) definește chiar repartiția normală de parametri m și σ;

2) $\alpha = \frac{1}{2}$ și $\beta = 0$ atunci (10.2.22) definește repartiția normală standard (normală de medie zero și dispersie unu).

Propoziția 10.2.11. Să presupunem acum un experiment pentru care cantitatea de informație $I(x)$ are forma

$$I(x) = \alpha x + \beta \ln x + \gamma,\ x > 0,\ \alpha > 0,\ \beta < 1\ \gamma \in \mathbf{R} \tag{10.2.23}$$

Atunci experimentul este deschis de repartiția gama generalizată

$$p(x; \alpha, \beta) = \frac{x^{-\beta} e^{-\alpha x}}{\alpha^{\beta - 1}\,\Gamma(1 - \beta)} \tag{10.2.24}$$

Scanned with OKEN Scanner

Demonstrație. Într-adevăr

$$1 = \int_0^\infty e^{-I(x)}\,dx \Rightarrow e^\gamma = \int_0^\infty x^{-\beta} e^{-\alpha x}\,dx \Rightarrow e^\gamma = \alpha^{\beta-1}\,\Gamma(1-\beta),$$ pentru

$\alpha > 0$, $\beta < 1$ și ținînd cont de (10.2.7) rezultă (10.2.24), care pentru $\alpha = 1$ definește repartiția gama simplă.

Consecință. Dacă $\alpha = \frac{1}{2\sigma^2}$ și $1 - \beta = \frac{\gamma}{2}$, $\gamma \in \mathbf{N}$ atunci cantitatea de informație (10.2.23) conduce la repartiția $H(m, \sigma)$ (χ^2_m de parametru σ).

Propoziția 10.2.12. Dacă

$$I(x) = \alpha \ln x + \beta \ln(1-x) + \gamma,\ x \in [0,1],\ \alpha < 1,\ \beta < 1,\ \gamma \in \mathbf{R} \tag{10.2.25}$$

atunci experimentul este descris de repartiția beta

$$p(x) = \frac{x^{-\alpha}(1-x)^{-\beta}}{B(1-\alpha, 1-\beta)}. \tag{10.2.26}$$

Demonstrație. Într-adevăr,

$$1 = \int_0^1 e^{-I(x)}\,dx \Rightarrow e^\gamma = \int_0^1 x^{-\alpha}(1-x)^{-\beta}\,dx = B(1-\alpha, 1-\beta)$$

pentru $\alpha < 1$, $\beta < 1$ și ținînd cont de (10.2.7) rezultă (10.2.26) remarcînd că pentru $\alpha = \beta = 0$ obținem repartiția uniformă pe $[0, 1]$.

Rezultatele de mai sus confirmă ideea posibilității determinării probabilității cu ajutorul informației, în ipoteza că fără a utiliza conceptul de probabilitate, am putea măsura într-un fel cantitatea de informație furnizată de un anume experiment. Ajungem astfel din nou la problema pusă la începutul lucrării, aceea a măsurării informației, acum însă cu condiția unei soluționări care să nu depindă de conceptul de probabilitate, problemă a cărei rezolvare ar conduce la o nouă formulare a teoriei informației, care ar putea sta la baza teoriei probabilităților, în calitate de concept primar.

Mai mult decît atît, rezultatele de mai sus ilustrează un exemplu teoretic de sistem cu conexiune inversă, sau cu feed back, în care intrarea și ieșirea sînt, pe rînd, conceptele de probabilitate și informație, aflate, după cum se poate constata de altfel, într-o relație generală de reciproc complet dependență.

Scanned with OKEN Scanner

Bibliografie

ACZÉL, J., *On different characterizations of entropies*, Lecture Notes in Math. 89 (1969), 1—11, Springer Verlag, Berlin.

ACZÉL, J., *Axiomatic characterization of some measures of divergence in information*. Z. Wahr. und Verw. Gebiete 21. (1972), 215 — 224.

ACZÉL, J., DAROCZY, Z., *Charakterisierung der Entropies positive Ordnung und der Shannonschen Entropie*. Acta Math. Acad. Sci. Hung., 14 (1963), 95 — 121.

BAD, M., BAD, E., *L'algorithme le plus rationnel de reconnaissance appliqué dans le diagnostique des maladies rénales*. La Santé Publique 1, 1972, 109 — 115.

BÂSCA, O., *Sisteme de stocare și regăsire a informației*. Teză de doctorat, Univ. București, 1977.

BEHARA, M., NATH, P., *Additive and non-additive entropies of finite measurable partitions*, Probability and Inf. Theory II, Lectures Notes, in Math., 296, 1973, 102 — 138, Springer Verlag, Berlin.

BELL, C. B., *Mutual information and maximal correlation as measures of dependence*. Ann. Math. Statist. 33, 1962, 587 — 595.

BÎRLEA, ST., *Aplicarea teoriei stărilor entropice la măsurarea gradului de organizare a sistemelor economice*, Probleme economice, 3, 1973.

BOORET, R., *A note on an information theoretic form of the uncertaintly principle*. Inf. and Control, 1, 1958, 398 — 401.

BOREL, E., *Probabilité et certitude*, Presses Univ. de France, 1969 .

BRILLOUIN, L., *Inevitable experimental errors, determinism and information theory*. Inf. and Control, 2, 1959, 45 — 63.

BRILLOUIN, L., *La science et la théorie de l'information*, Paris, 1959.

CAMPBELL, L. L., *A coding theorem and Rényi's entropy*. Inf. Contr. 8, 1965, 423 — 429.

CAMPBELL, L. L., *Entropy as a measure*. IEEE Trans. Inform. Theory, IT-11, 1965, 112 — 114.

CAMPBELL L. L., *Characterization of entropy of probability distributions on the real line*. Inf. Contr. 21, 1972, 329 — 343.

CERNY, J., BRUNOWSKY, P., *A note on information without probability*. Inf. and Control, 25, 1974, 134 — 144.

CIUCU, G., IOSIFESCU, M., THEODORESCU, R. *Teoria jocurilor*. Ed. tehnică, București, 1965.

CIUCU, G., CRAIU V., *Inferența statistică*. Ed. Did. și Ped., București, 1974.

Scanned with OKEN Scanner

CONSTANTINESCU, I., CONDREA, D., NICOLAU, E., *Teoria Informației*. Ed. Tehnică, București, 1958.
CRAIU, V., *Verificarea ipotezelor statistice*. Ed. Ped. și Did,. București, 1972.
CUCULESCU, I., *Teoria probabilităților*, Univ. București, 1976.
CSISZAR, I., *Information-type measure of difference of probability distributions and indirect observations*. Studia Sci. Math. Hung., 2, 1967, 299 — — 318.
CSISZAR, I., *On topological properties of f-divergence*. Studia Sci. Math. Hung. 2, 1967, 329 — 339.
CYRANSKI, F. J., *Measurement, Theory and Information*. Inf. and Control, 41, 1979, 275 — 304.
DAROCZY, Z., *Über die gemeinsame Charakterisierung der zu den nicht vollständigen Verteilungen gehörigen Entropien von Shannon und von Rényi*. Z. Wahrsch. 1, 1963, 381 — 388.
DAROCZY, Z., *Über Mittelwerte und Entropien volls*. Wahr. Acta Math. Acad. Sci. Hung., 15, 1964, 203 — 210.
DINESCU, C., SĂVULESCU, B., VASILIU, D. P., *Metode matematice pentru fundamentarea deciziilor în producție*. Edit. tehnică, București 1986.
DUTTA, M., *On maximum (information-theoretic) entropy estimation*. Sankhya 28, 1966, 319 — 328.
ERDÖS, P., *On the distributions function of additive functions*. Ann of Math. 47, 1946, 1 — 20.
FADEEV, D. K., *Despre noțiunea de entropie a unei scheme probabilistice finite*. În culegerea „Probleme de cibernetică" I.S.R.S., 1959, 514 — 520.
FANO, R. M., *Transmission of information*. New York, 1961.
FEINSTEIN, A. *Foundations of information theory*, New York, 1958.
FEINSTEIN, A., *On the coding theorem and its converce for finite memory channels*. Information and Control, 1, 1959, 25 — 44.
FERON, R., *Information et régression*. C.R. Acad. Sci. Paris, 232, 1951, 1636—1638.
FERON, R., *Information et correlation*. C. R. Acad. Sci. Paris, 234, 1952, 1343—1345.
FISHER, R. A., *Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population*. Biometrika, 10, 1915, 507 — 521.
FISHER, R. A., *Probability, likelihood and quantity of information in the logic of uncertain inference*. Proc. Royal Soc. A. 146, 1934, 1 — 8.
FORTE, B., DAROCZY, Z. *A characterization of Shannon's entropy*. Boll. Un. Mat. Ital., 4 — 5, 1968, 631 — 635.
FRASER, D. A. S., *On information in statistics*. Ann. Math. Statist. 36, 1965, 890 — 896.
GEORGESCU,ROEGEN, N., *Legea entropiei și procesul economic*. Ed. Politică, București, 1979.
GIBBS, J. W., *Elementary principles in statistical mechanics developed with especial reference to the rational foundation of thermodynamics*. New Haven, 1902.
GOLDMAN, S., *Information theory*, Londra, 1953.
GUIAȘU, S., *Aplicații ale teoriei informației. Sisteme dinamice. Sisteme cibernetice*. Ed. Acad. RSR, București, 1968.
GUIAȘU, S., *Information theory with applications*. MgGraw-Hill New York — Düsseldorf, 1977.
GUIAȘU, S., MALITA M. *Triade*. Ed. Științifică, București, 1973.
GUIAȘU, S., THEODORESCU, R. *Teoria matematică a informației*. Ed. Acad. RSR, București, 1966.

Scanned with OKEN Scanner

GUIAȘU, S., THEODORESCU, R. *Incertitude et information*, Les Presses de l'Université Laval, Québec, 1971.
HAJEK, J., *A property of J-divergence of marginal probabilistic distributions*. Czechosl. Math. Journal 8(83), 1958, 460 — 463.
HAMDAN, M. A., TSOKOS, C. P., *An information measures of association in contingency tables*. Inf. and Control, 19, 1971, 174 — 180.
HARTLEY, R. V. H., *Transmission of Information*. Bell System Technical Journal 3, 1928, 535 — 563.
HATORI, H., *A note on the entropy of a continuous distribution*. Kodai Math. Sem. Rep., 10, 1958, 172 — 176.
HINCIN. A. I., *Noțiunea de entropie în calculul probabilităților*. Analele româno-sovietice 2, 1954, seria mat.— fiz., 23 — 49.
HINCIN, A. I., *Teoremele fundamentale ale teoriei informației*. În culegerea „Probleme de cibernetică", I.S.R.S., 1959, 440 — 513.
HORIBE, Y., *A note on entropy metrics*. Inf. and Control 22, 1973, 403 — 404.
IAGLOM, A. M., IAGLOM, I. M., *Probabilitate și informație*, Ed. Did. și Ped., București, 1963.
IKEDA, S., *An characterization of the Kullback-Leibler mean information for continuous probability distributions*. Ann. Inst. Statist. Math. 14, 1962, 73 — 79.
INGARDEN, R. S., *Information theory and variational principles in statistical theories*. Bull. Acad. Polon. Sci. Sci. Math. Astron, Phys., 11, 1963, 541 — 547.
INGARDEN, R. S., *Simplified axioms for information without probability*. Prace Matematyczne 9, 1965, 273 — 285.
INGARDEN, R. S., *Information theory and Termodynamics*. Pergamon Press, Oxford, 1977.
INGARDEN, R. S., KOSSAKOWSKI, A., *Poisson probability distribution and information thermodynamics*. Bull. Acad. Polon. Sci. Ser. Math. Astronom. Phys. 19, 1971, 83 — 86.
INGARDEN, R. S., URBANIK, K., *Information without probability*. Col., Math. 9, 1962, 281 — 304.
IOSIFESCU, M., *Sampling entropy for random homogeneous systems with complete connections*. Ann. Math. Statist. 36, 1965, 1433 — 1436.
IOSIFESCU, M., MIHOC, GH., THEODORESCU, R., *Teoria probabilităților și statistica matematică*, Ed. Tehnică, București, 1966.
ITO, H., *Principle of minimum entropy in information theory*. Proc. Japan Acad., 29, 1953, 194 — 197.
JAYNES, E. T., *Information theory and statistical mechanics, I, II*, Phys. Rev. 106, 1957, 620 — 630, 108, 1957, 171 — 190.
KAMPÉ DE FÉRIET, J., *Théorie de l'information. Principe du maximum de l'entropie et ses applications à la statistique et à la méchanique*. Publications du Laboratoire de calcul de la Fac. de Sci de l'Univ. de Lille, 1963.
KAMPÉ DE FÉRIET, J., FORTE, B., *Information et probabilité*. C. R. Acad. Sci. Paris, 265, séries A et B, 1967, 110 — 114, 142 — 146, 350 — 353.
KAPUR, J. N., *Generalized entropy of order α and type β*. Math. Seminar, Delhi, 4, 1967, 78 — 94.
KAPUR, J. N., *On information of order α and type β*. Proceedings Indian Academy of Sciences, 68 A, 1968, 65 — 75.
KENDALL, D. G., *Functional equations in information theory*. Z. Wahrscheinlichkeitstheorie, 2, 1964, 225 — 229.
KENDALL, M. G., *Rank correlation methods*, London, 1948.
KOGAN, I. M., *Prikladnaia teoria informații*, Moskva, 1981.
KOLMOGOROV, A. N., *Cibernetica*. Marea Enciclopedie Sovietică, 51, 1958, 149 — 151

Scanned with OKEN Scanner

TAKANO, K., *On the basic theorems of information theory.* Ann. Inst. Statist. Math, 2, 1958, 53 — 77.
TÖVISSI, L., THEILER, G., *Entropia, eventual alte măsuri cu suport probabilistic în evaluarea unor informații economice".* Rev. de Stat, 1, 1977.
TVERBERG, H., *A new derivation of the information function.* Math. Scand., 6, 1958, 297—298
URBANIK, K., *On the definition of information.* Reports on Math. Phys., 4, 1973, 289 — 301.
VARMA, R. S., *Generalisations of Rényi's entropy of order α.* J. Math. Sci. Delhi, 1, 1966, 34 — 38.
VADUVA, I., *Analiza dispersională.* Ed. Tehnică, București. 1970.
VITUSHKIN, A. G., *Theory of Transmission and Processing of Information,* Pergamon Press, Oxford (1961)
VON NEUMANN, J., *The general and logical theory of automata.* Massachussets, Inst. of Tech., 1948.
VON NEUMANN, J., MORGENSTERN, O., *Theory of games and economic behaviour.* Second edition, Princeton University Press, Princeton, New York, 1947.
WATANABE, S., *A note on the formation of concept and of association by information-theoretical correlation analyses.* Inf. and Control, 4. 1961, 291 — 296.
WATANABE, S., *Methodologies of pathern recognition.* Academic Press, New York-London, 1969.
WATANABE, S., *Knowing and guessing.* A quantitative study of inference and information. Wiley, New York-London-Sydney-Toronto, 1969.
WIENER, N., *Cybernetics.* Hermann, Paris, 1948.
WIENER, N., *Cibernetica sau știința comenzii și comunicării la ființe și mașini.* Ed. Științifică, București, 1966.
WOLFOWITZ, J., *Coding theorems of information theory.* Berlin, 1969, ed. II.
YULE, G. U., *On the theory of correlation.* J.R.S.S. A., 60, 1897, 812 — 854.
YULE, G. U., KENDALL, M. C., *Introducere în teoria statisticii.* Ed. St., București, 1969.
ZADEH, L. A., POLAK, E., ș.a. *Teoria sistemelor.* Ed. Tehnică, București, 1969.

Scanned with OKEN Scanner

Cuprins

I. Măsura informației 7
1. Apariția teoriei informației 7
2. Problema măsurării informației 12
2.1. Considerații generale 12
2.2. Axiomele măsurii cantității de informație 14
3. Entropia Shannon 17
3.1. Entropia unei variabile aleatoare discrete 17
3.2. Entropia unei variabile continue 20
3.2.1. Entropia Boltzmann-Shannon-Wiener 20
3.2.2. Entropia ca variație a informației 28
3.3. Entropii ale unui vector aleator discret 33
3.3.1. Cazul bidimensional 33
3.3.2. Cazul tridimensional 40
3.3.3. Cazul n-dimensional 47
3.4. Entropii ale unui vector aleator continuu 53
3.5. Unicitatea entropiei Shannon. Extensiuni 57
3.5.1. Axiomatica lui C. E. Shannon (1948) 57
3.5.2. Axiomatica lui A. I. Hincin (1956) 58
3.5.3. Axiomatica lui D. K. Fadeev (1956) 59
3.5.4. Axiomatica lui H. Halori (1958) 60
3.5.5. Epsilon-entropia lui A. N. Kolmogorov (1958) . . . 61
3.5.6. Axiomatica și alfa- entropia lui A. Rényi (1961) . . . 63
3.5.7. Axiomatica R. S. Ingarden și K. Urbanik (1962) . . 67
3.5.8. Entropia Onicescu (1963) 68
3.5.9. Entropia Rényi d-dimensională de ordin α (1970) . . . 69
3.5.10. Alfa-entropia Behara-Nath (1973) 71
4. O măsură a utilității informației 75
4.1. Utilitate 75
4.1.1. Axiomatica Von Neumann-Morgenstern (1944) . . . 76
4.1.2. Axiomatică Iuce-Raiffa (1957) 78

Scanned with OKEN Scanner

4.2. Entropia ponderată a lui S. Guiașu (1968) 80

II. MĂSURA CORELAȚIEI

5. **Problema măsurării corelației** 87
5.1. Considerații generale 87
5.2. Postulatele lui A. Rényi (1959) 88
6. **Măsuri clasice** 89
6.1. Corelația Galton-Pearson (1901) 89
6.1.1. Coeficient de corelație simplă 90
6.1.2. Rapoarte de corelație 93
6.1.3. Coeficient de corelație multiplă 98
6.1.4. Coeficient de corelație parțială 100
6.2. Măsuri ale concordanței 102
6.2.1. Coeficientul lui G. Fechner (1897) 102
6.2.2. Coeficientul lui G. Spearman (1904) 103
6.2.3. Coeficientul lui M. G. Kendall (1948) 104
6.3. Media pătratului de contingență a lui K. Pearson (1901) . . 105
6.4. Corelația maximală Gebelein-Kramer (1941) 108
6.5. Coeficientul asimptotic de corelație Onicescu-Săcuiu (1964) . 115
6.6. Coeficientul de corelație globală a lui I. Săcuiu (1970) . . . 116

7. **Măsuri informaționale** 118
7.1. Coeficientul logaritmic de corelație M. C. Camargo—I. Medina (1955) 118
7.2. Coeficientul de corelație informațională al lui E. H. Linfoot (1957) 120
7.3. J-divergența lui S. Kullback (1959) 122
7.4. Coeficientul de corelație informațională a lui C. B. Bell (1962) 133
7.5. Energia și corelația informațională a lui O. Onicescu (1966) . 135
7.6. Corelație și coeficienți de corelație entropică (1969, 1973, 1976) 138
7.6.1. Măsuri entropice ale dependenței globale 138
7.6.2. Măsuri entropice ale dependenței unilaterale 151
7.6.3. Măsuri entropice ponderate ale dependenței 154
7.6.4. Conexiuni ale măsurilor entropice cu alte măsuri ale dependenței 160

III. INFORMAȚIE ȘI CORELAȚIE ENTROPICĂ ÎN TEORIA SISTEMELOR 165
8. **Interdependența** 165
8.1. Sisteme 165
8.1.1. Sisteme de transmitere a informației 167

Scanned with OKEN Scanner

8.1.2. Sisteme de strategii alealoare 174
8.2. Măsuri entropice ale interdependenței 179
8.3. Metrici entropice 186
9. Organizare 191
9.1. Măsuri entropice ale organizării 191
9.2. Clasificare entropică 201
10. Informație și probabilitate 206
10.1. Principiul informației maxime 206
10.2. Conexiunea dintre informație și probabilitate 216
Bibliografie 227

Scanned with OKEN Scanner

Redactor: ION MIHAI
Tehnoredactor: MARILENA DAMASCHINOPOL

Coli de tipar: 15, Bun de tipar: 18.04.1988

c. 425 I. P. INFORMAȚIA,
str. Brezoianu, nr. 23—25,
București

Scanned with OKEN Scanner